Development of new fault detection methods for rotating machines (roller bearings)

POLITECNICO DI TORINO

Mechanical and Aerospace Engineering Department

A Thesis submitted to Politenico di Torino
for the degree of Doctor of Philosophy

Development of new fault detection methods of rotating machines (rolling bearing)

Ali akbar
Tabrizi zarringhabaei

Supervisor

Prof. Luigi Garibaldi

March 2015

Acknowledgement

First of all, I would like to express my gratitude to my supervisor, Prof. Luigi Garibaldi for all his help, guidance, encouragement and knowledge. It was an honour to work with him. I would like to thank him for being a friend to me apart from a supervisor. I would also like to thank Prof. Irina Trendafilova for accepting me in the Dynamics of Advanced Structures and Machinery (DASM) research group in University of Strathclyde, and for her valuable advice and support. Thanks to Prof. Alessandro Fasana, Stefano Marchesiello and other colleagues in the Dynamics & Identification Research Group (DIRG) in Politecnico di Torino and Hussein Al-bugharbee in DASM research group in Strathclyde University for their valuable discussions and help for providing the experimental data.

Finally, my special appreciation goes to my wife Maryam and my dear son Aidin since this work was not possible without their support, care, love and patience.

Abstract

Early fault diagnosis of roller bearings is extremely important for rotating machines, especially for high speed, automatic and precise machines. Many research efforts have been focused on fault diagnosis and detection of roller bearings, since they constitute one the most important elements of rotating machinery. In this study a combination method is proposed for early damage detection of roller bearing. Wavelet packet transform (WPT) is applied to the collected data for denoising and the resulting clean data are break-down into some elementary components called Intrinsic mode functions (IMFs) using Ensemble empirical mode decomposition (EEMD) method. The normalized energy of three first IMFs are used as input for Support vector machine (SVM) to recognize whether signals are sorting out from healthy or faulty bearings. Then, since there is no robust guide to determine amplitude of added noise in EEMD technique, a new Performance improved EEMD (PIEEMD) is proposed to determine the appropriate value of added noise. A novel feature extraction method is also proposed for detecting small size defect using Teager-Kaiser energy operator (TKEO). TKEO is applied to IMFs obtained to create new feature vectors as input data for one-class SVM. The results of applying the method to acceleration signals collected from an experimental bearing test rig demonstrated that the method can be successfully used for early damage detection of roller bearings.

Most of the diagnostic methods that have been developed up to now can be applied for the case stationary working conditions only (constant speed and load). However, bearings often work at time-varying conditions such as wind turbine supporting bearings, mining excavator bearings, vehicles, robots and all processes with run-up and run-down transients. Damage identification for bearings working under non-stationary operating conditions, especially for early/small defects, requires the use of appropriate techniques, which are generally different from those used for the case of stationary conditions, in order to extract fault-sensitive features which are at the same time insensitive to operational condition variations. Some methods have been proposed for damage detection of bearings working under time-varying speed conditions. However, their application might increase the instrumentation cost because of providing a phase reference signal. Furthermore, some methods such as order tracking methods still can be applied when the speed variation is limited. In this study, a novel combined method based on cointegration is proposed for the development of fault features which are sensitive to the presence of defects while in the same time they are insensitive to changes in the operational conditions. It does not require any additional measurements and can identify defects even for considerable speed variations. The signals acquired during run-up condition are decomposed into IMFs using the performance improved EEMD method. Then, the cointegration method is applied to the intrinsic mode functions to extract stationary residuals. The feature vectors are created by applying the Teager-Kaiser energy operator to the obtained stationary residuals. Finally, the feature vectors of the healthy bearing signals are utilized to construct a separating hyperplane using one-class support vector machine. Eventually the proposed method was applied to vibration signals measured on an experimental bearing test rig. The results verified that the method can successfully distinguish between healthy and faulty bearings even if the shaft speed changes dramatically.

Table of contents

Chapter 1

Introduction

1.1 Rotating machines

Rotating machineries are the most common mechanical equipment in industrial application; their main components are gearboxes, roller bearings and rotary shafts. Since high-quality, low-cost products and safe production are highly requested in modern rotating machines, high precision and high automation levels are simultaneously demanded for their characteristic. However, they generally works under severe operating conditions, which may cause machinery to break down and decrease their performance which in turns means lower products quality, less operation safety and so on. Therefore, in addition to their reliability, the capability of detecting possible faults at a very early stage represents a fundamental task.

Roller bearings, illustrated in Fig. 1.1, are widely used in rotating machinery and one of the major reasons for machine breakdown is their failure. They are often working under harsh industrial conditions and where basic dynamic loads of machines are applied. As a finite number of rolling elements have to carry the load, rolling element bearings generate vibrations even if they are geometrically perfect. Vibration based condition monitoring is the most common method for extracting some important information to identify the state of machines and probable defects. The concept behind vibration analysis is that the vibration signature of the component examined changes in the case of damage occurrence. When machines are in healthy conditions, the vibrations generated are linked to their shaft rotation, the gear meshing of the gear stages of their gearbox, and other periodic or non-periodic events occurring while in operation.

In this chapter the typical damages of bearings are introduced. Then, in Section 1.3 and 1.4 two different approaches to diagnostics of machines, model based or data based, are described, respectively. Section 1.3 focuses on the main steps of the data based methods, which include data acquisition, pre-processing, feature extraction, post-processing and pattern recognition.

1.2 Localized and distributed damage

Damage is interpreted as changes introduced into a system that affect the current performance of the system [1]. The most common types of faults happening in rotating machinery are shaft faults, gear faults and bearing. Shaft faults include unbalance, misalignment, shaft bending, whirl and cracking. Some typical gear faults are broken tooth, cracks, spalls, wearing and pitting. Bearing defects that some of them are shown in Fig. 1.2 can be considered as belonging to two different classes; distributed and local damages. The distributed damages are characterised by regularly spaced indentations on the entire raceway circumference and can be noted as surface roughness, waviness, abrasive wear and brinelling. They are usually due to manufacturing error, overloading, misalignment, foreign particle contamination, etc. Brinelling shown in Fig. 1.2b is depression created on the contact surface. It is also small depression on the rolling surface caused by solid foreign matters. The localised defects include cracks, corrosion, identation, flaking, pitting and spalling. Indentation shown in Fig. 1.2d is a defect caused by foreign particles. In flaking (Fig. 1.2e), the bearing surface turns scaly and peels off due to contact load repeatedly applied to the raceway and rolling surface. Pittings are small holes (0.1 mm in depth) generated on the raceway surface by rolling fatigue. However, electric pittings seen in Fig. 1.2f are created by sparks generated when electric current passes through a thin oil film at the rolling contact point. Spalling seen in Fig. 1.2g appears when a fatigue crack begins towards the surface and a piece of metal breaks away and creates a small pit or spall.

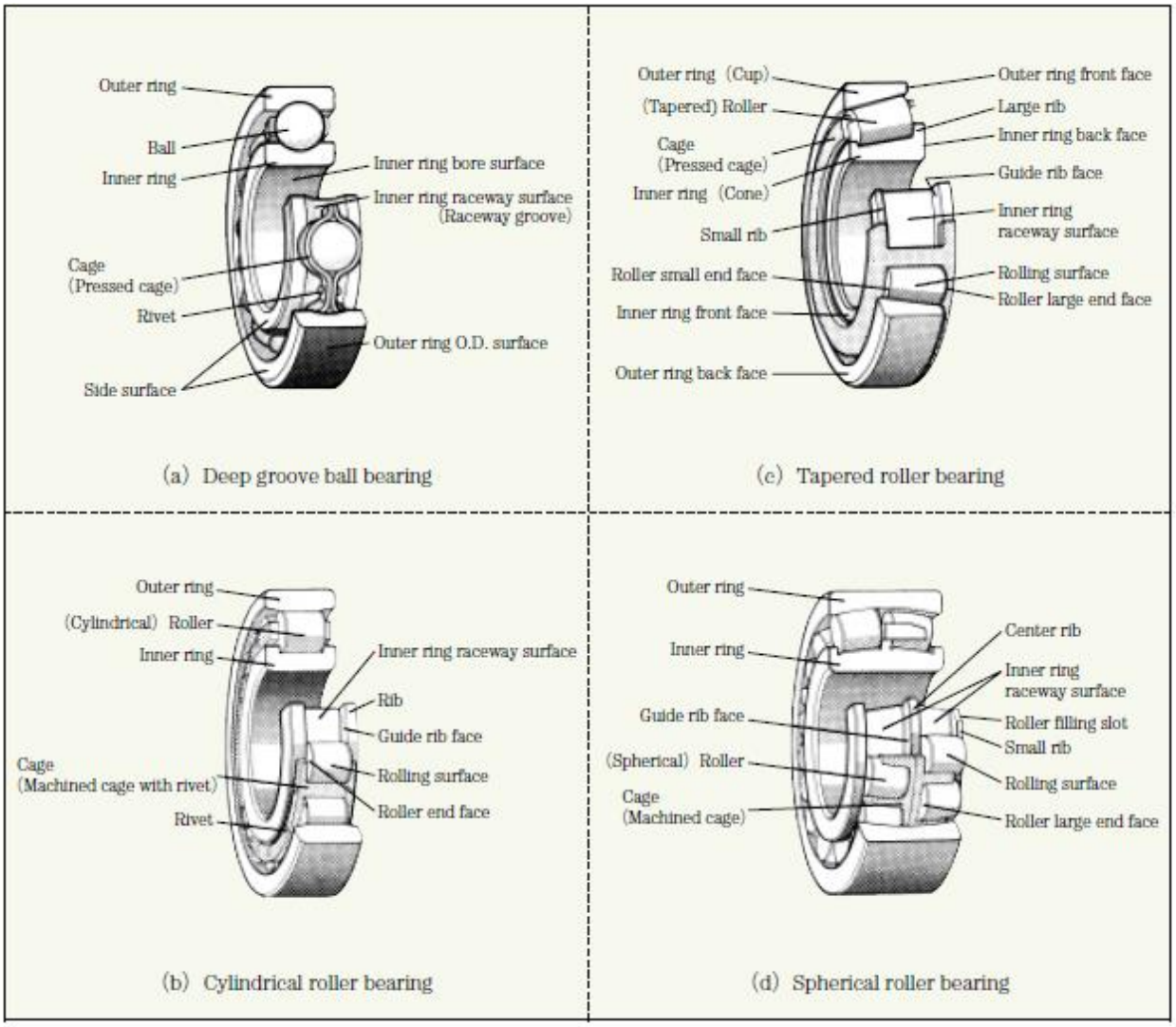


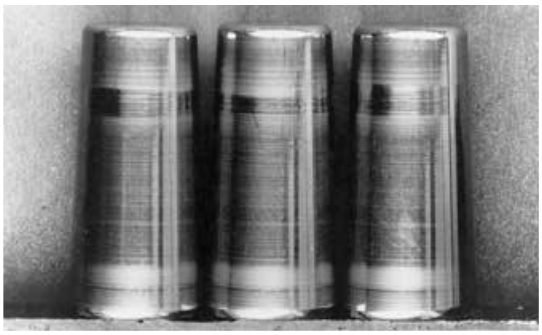
Figure 1.1- Rolling element bearings and description of each part (from [2]).

An exhaustive description of all possible bearing failures, with respective causes and countermeasures, is presented in [5].

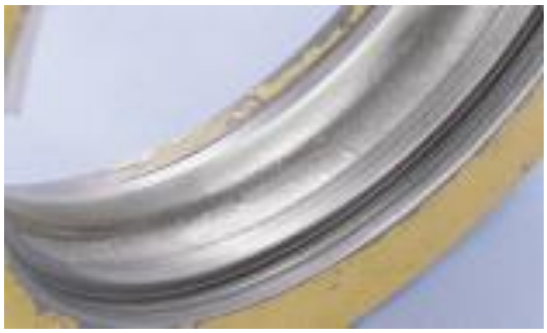
The vibration signals generated by faults are often used for damage detection, since they often carry some significant dynamic information. In order to implement the vibration based analysis, two methods have been adopted for creating the localized defects on bearings. Running the bearing until failure [6-11] or creating defects in the bearings by techniques such as acid etching, spark erosion, scratching or mechanical indentation [12-20]. In the first approach, harsh conditions are applied to bearings like overloading, over speeding or lack of enough lubricants.

1.3 Data based diagnostic

There exist two approaches for diagnostic: Data based diagnostic and Model based. In the data based approach, informative data is collected from the system to implement appropriate analysis, whereas the model based builds a explicit mathematical model based on the dynamics of the of the



(a) (from [3])



(b) (from [4])

(c) (from [4])

(d) (from [4])

(e) (from [2])

(f) [from [2])

(g) (from [4])

Figure 1.2- Some of bearing damages: (a) wear (b) brinelling (c) corrosion (d) identation (e) flaking (f) electric pitting (g) spalling

Figure 1.3- The procedure of the data based diagnostic method.

monitored machine. The procedure of the data based method is shown in Fig. 1.3, which consists of data acquisition, pre-processing, feature extraction, post-processing and pattern recognition.

1.3.1 Data acquisition

The first demand of an intelligent fault detection system is that it should measure the appropriate data. It deals with the type of data, the sensor type, number and location and the data acquisition system, as well as all the acquisition parameter, such as sampling frequency, filtering, etc. The measured data used for identifying the condition of the system can be vibration data, acoustic data, oil analysis data, temperature, pressure, humidity, weather or environment data, etc. Various sensors, such as accelerometers, velocity transducers, dual vibration probes, laser vibrometers, ultrasonic sensors, acoustic emission sensors, encoders, tachometers, etc., have been provided to measure different types of signals [21]. Different techniques for multiple sensor data fusion are also discussed in [22].

Data acquisition transducers and recording equipment used to monitor are discussed in detail in [23-28]. The appropriate type and location of transducers depends upon the type of machinery and its construction. The suitable placement of transducers is discussed in [26].

1.3.2 Pre-processing

As the measurement instruments are not perfect, the collected physical information will be polluted by different type of noise (e.g. the quantisation noise unavoidable in adopting a digital representation, the sensor response, the measurement noise, the unrelated noise from other sources). In order to prepare the data for feature extraction, cleansing the data is definitely required such as in [1]: filtering to remove noise, spike removal by median filtering, removal of outliers, and treatment of missing data values. To avoid the risk of removing some information, cleansing must be carefully performed. The dimension reduction of the data might be another purpose of the pre-processing as well as its transformation. An example could be to transfer the data from time domain to a spectrum, for a gearbox. The windowing and averaging process for this step would result in far fewer spectral lines than the original time-series points. However, this is not an automatic step and often is carried out based on experience and engineering judgment.

There exist some different approaches for noise cancelling such as classical approach (e.g. Kalman filtering), adaptive filtering and Wavelet transform (WT) approaches. The Kalman filter is an optimal estimator used for linear systems. Adaptive filtering try to separate two uncorrelated components (e.g. deterministic and random components) using a reference signal. The WT approach decomposes the signal through highpass and lowpass filters. Then, the predetrmined threshold is applied so that the impulses can be effectively retained.

Filters utilized to remove noise can be linear or non-linear. The linear filters can be classified into a number of different band-forms describing which frequency bands the filter passes (the pass-band) and which it rejects (the stop-band):

- Low-pass filter: low frequencies are passed, high frequencies are rejected.
- High-pass filter: high frequencies are passed, low frequencies are rejected.
- Band-pass filter : only frequencies in a frequency band are passed.
- Band-reject (Notch) filter: often used to remove harmonics given by gearmesh.

A non-linear filter is a filter whose output is not a linear function of its input. The linear filters are often used to remove noise and distortion that was created by nonlinear processes, simply because the proper non-linear filter would be too hard to design and construct.

1.3.3 Feature extraction (signal processing)

This section contains an overview of the most important signal processing methods for condition monitoring over the years, summarised in Fig. 1.4 [29]. The main discrimination is whether the methods are appropriate for stationary or non-stationary signals. Non-stationarity can be the consequence of time-varying operation or environment of the structure or time-varying dynamics (due to the evolution of damage or varying boundary conditions).

A short history of bearing diagnostics was published in [30]. Balderston [31] proposed one of the earliest papers on bearing diagnostics in 1969. He found that the signals generated by bearing faults were located in the high frequency region of resonances excited by the internal impacts. In the following years, the synchronous averaging was used to expose local faults in both bearings and gears [32]. A fundamental analysis of synchronous averaging was presented by Braun [33] in 1975. He applied the basic technique to bearing signals and notice that bearing signals are not completely periodic [34]. At around that time, in order to shift the frequency analysis from the very high range of resonant carrier frequencies to the much lower range of the fault frequencies, the High frequency resonance technique (HFRT) was developed, so that they could be analysed with good resolution [35]. The first studies on vibration monitoring techniques were proposed in the 1980s. Mathew and Alfredson presented a brief review of time and frequency domains methods [36]. McFadden and Smith investigated the monitoring by the high-frequency resonance technique [37]. A detailed review of the different vibration and acoustic methods, such as vibration measurements in time and frequency domains, sound measurements, the shock pulse method and the acoustic emission technique were proposed by Tandon and Nakra [38, 39].

Many reviews on bearing diagnostics techniques have been proposed, such as [22, 30, 40-42].

1.3.3.1 Time domain

Since the time domain techniques are directly based on the time waveform, they are considered among the first diagnostic. There exist some simple statistical measures to compare and identify the state of the machine such as the peak-to-peak, overall root mean square (RMS) level and the crest factor, which often are not so sensitive to small defects. They are expressed as following:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{1.1}$$

$$Crest\ factor = \frac{max\left(abs(x)\right)}{RMS} \tag{1.2}$$

where x is the collected signal and N is the number of its elements.

The centred moments of a signal is defined as:

$$C_n = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^n \tag{1.3}$$

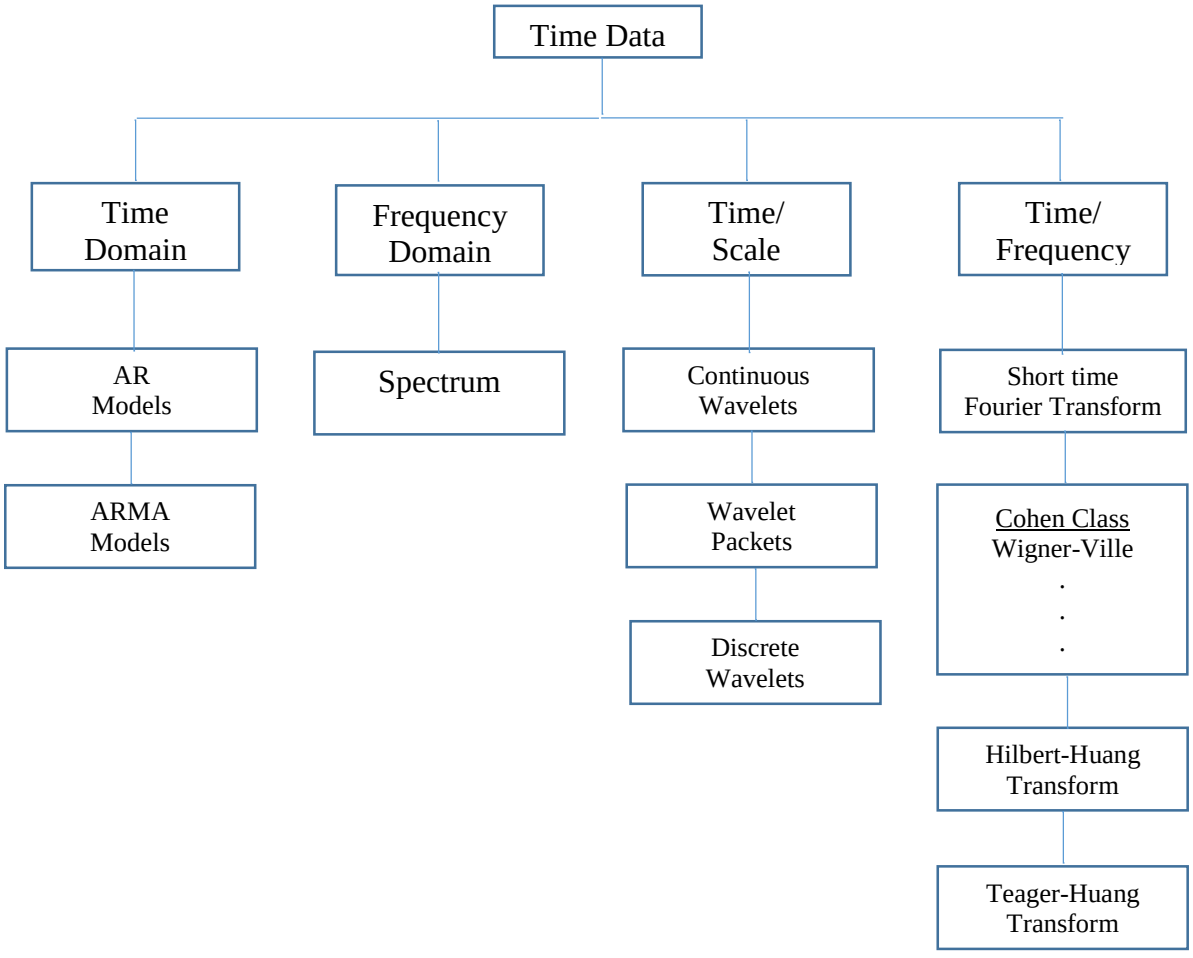


Figure 1.4- Methods of feature extraction (signal processing).

The normalized third and fourth moments called skewness and kurtosis, respectively, are quite useful in defect detection and calculated as:

$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{\frac{3}{2}}} \tag{1.4}$$

$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{2}} \tag{1.5}$$

The kurtosis has been common for condition monitoring applications [43, 44]. The spectral kurtosis is a recent interesting development in condition monitoring [30].

Another approach related to the time domain analysis is the application of time series models to waveform data [22]. The main idea is to fit the waveform data to a parametric time series model and to extract features based on this parametric model. Two examples are the simplest so called autoregressive (AR) model and the autoregressive moving average with eXogenous inputs (ARMAX) model expressed as Eqs. 1.6 and 1.7, respectively:

$$y_i = \sum_{j=1}^{p} a_j\, y_{i-j} + \varepsilon_i \tag{1.6}$$

where p denotes the model order and ε_i is unmeasured Gaussian white noise.

$$y_i = \sum_{j=1}^{p} a_j\, y_{i-j} + \sum_{j=1}^{q} b_j\, x_{i-j} + \sum_{j=1}^{r} c_j\, \varepsilon_{i-j} \tag{1.7}$$

The parameter estimation procedures for these models can be found, in details, in [45]. The residuals (prediction error) is another feature created by the time-series models. When a good time-series model is created in healthy condition, it will make good predictions, i.e. the sequence of prediction errors (residuals) between model outputs and measured data have low variance. If the system changes to a damaged condition, the model will not make good predictions and the variance of the residual will be high; a good recent application is given in [46].

1.3.3.2 Frequency domain

A common way of implementing the feature extraction is to make a linear or nonlinear transform. This is often carried out by using Fourier analysis. The idea is that the function of interest is represented as a sum over a set of fixed basis functions. The Fourier representation

provides the most popular means of visualisation of the signal characteristics, through the spectrum. This is usually defined via the Fourier transform for a continuous time signal.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)\, e^{-i\omega t}\, dt \tag{1.8}$$

$$X_i = \sum_{j=1}^{N} x_i\, e^{-ij\omega\Delta t} \tag{1.9}$$

The application of Fourier analysis has been widely used because of the efficiency of the Fast Fourier Transform algorithm (FFT) [47].

The frequency-domain approaches rely on the detection of the characteristic rotational frequencies related to specific bearing element faults. These frequencies could be noticed by observing the envelope of the vibration signal acquired for a damaged bearing. They are as following:

Ballpass frequency, outer race:

$$BPFO = \frac{nf_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{1.10}$$

Ballpass frequency, inner race:

$$BPFI = \frac{nf_r}{2}\left\{1 + \frac{d}{D}\cos\phi\right\} \tag{1.11}$$

Fundamental train frequency (cage speed):

$$FTF = \frac{f_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\} \tag{1.12}$$

Ball (roller) spin frequency:

$$BSF(RSF) = \frac{D}{2d}\left\{1 - \left(\frac{d}{D}\cos\phi\right)^2\right\} \tag{1.13}$$

where n is the number of rolling elements, f_r the shaft speed, d the diameter of the rolling element, D the pitch diameter and ϕ the contact angle (Fig. 1.5).

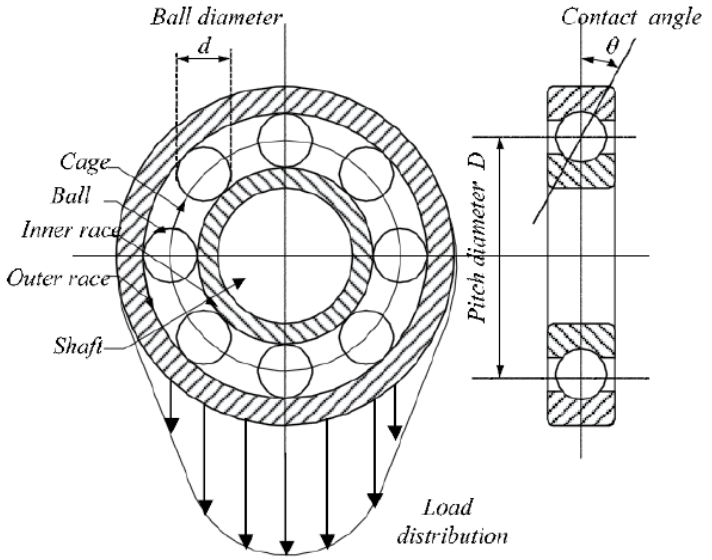


Figure 1.5- Rolling element bearing components and the load distribution (from [48]).

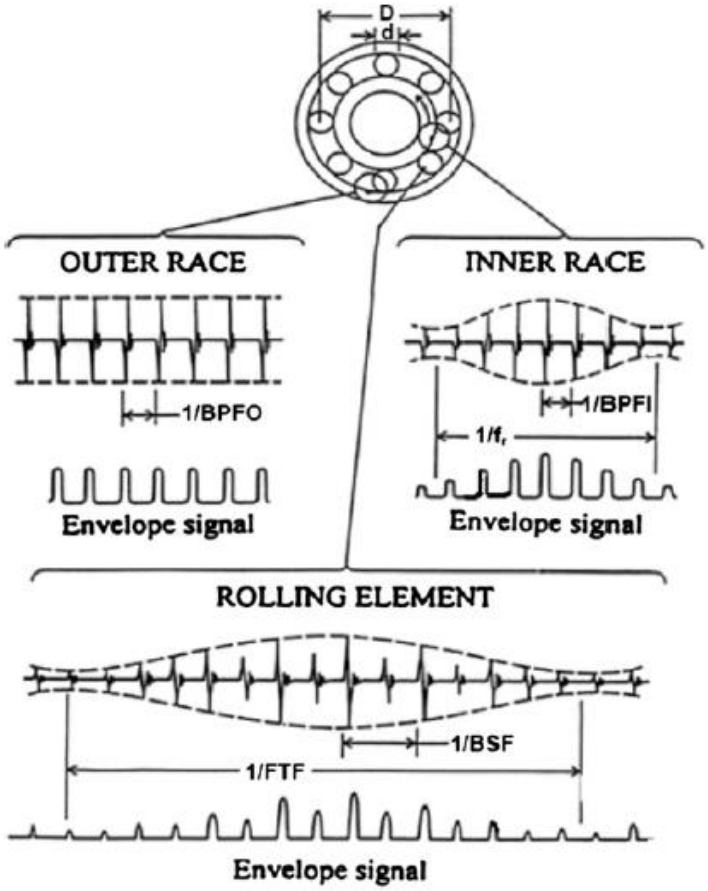


Figure 1.6- Typical signal and envelope signals from local faults in rolling element bearing (from [30])

An example is in Fig. 1.6, where typical acceleration signals produced by localised faults in the various components of a rolling element bearing, and the corresponding envelope signals produced by amplitude demodulation, are presented. It illustrates typical modulation patterns for unidirectional (vertical) load on the bearing, at shaft speed for inner race faults, and cage speed for rolling element faults.

Another ancient technique, which is able to detect harmonics and sideband patterns in power spectrum is the cepstrum first presented in [49]. It was defined as the power spectrum of the logarithmic power spectrum. Later, it has been redefined as the inverse Fourier transform of the logarithmic spectrum:

$$C(\tau) = F^{-1}\left[log\left(X(f)\right)\right] \tag{1.14}$$

where τ has the dimensions of time but is named quefrency and:

$$X(f) = F[x(t)] = A(f)\, e^{j\phi(f)} \tag{1.15}$$

$$log\left(X(f)\right) = ln\left(A(f)\right) + j\phi(f) \tag{1.16}$$

The power cepstrum of a signal, which is the most common among several version of cepstrum is defined as the squared magnitude of the inverse Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal (Eq. 1.17).

$$power\ cepstrum\ [X(f)] = |F[log(|F[x(t)]|^2)]|^2 \tag{1.17}$$

Some applications of the vibration spectrum and cepstrum are given in references [50-52].

The frequency domain methods are certainly among the most common feature extraction techniques for bearing fault detection due to its ability in identifying and isolating certain frequency components of interest. However, the fundamental assumption is that the analysed data are stationary or can be reduced to stationarity by a simple transformation. If the frequency content of a signal does vary with time, Fourier analysis will provide a time-averaged summary. Thus, it is not an appropriate method for non-linear, non-stationary signals such as the case of nonstationary running conditions, rolling elements sliding, rotating fluctuations and so on.

1.3.3.3 Time–Frequency/Scale/Order Domain

Most of the traditional signal processing methods can only be applied to stationary signals and cannot reveal the local features in both time and frequency domains simultaneously. The time-frequency techniques are powerful methods to identify the health information and extract the time variant features and the frequency components from acquired nonstationary signals. There exist various time-frequency/scale/order methods applied to identify faults of rotating machines such as Short time Fourier transform, wavelet transform (WT), Wigner-Ville distribution (WVD), Hilbert-Huang transform (HHT) and Teager-Huang transform (THT). A comprehensive information can be found in the review presented by Feng et al. [53].

Short time Fourier transform

In order to reveal the time-varying nature of a signal, a short time window slides along time to break down the signal into a number of small segments, which can be considered stationary in a short duration. Then, the local Fourier spectrum of each segment around the time location of the short time window is obtained [54].

For the signal $x(\tau)$, the short time Fourier transform is defined as follows:

$$STFT_x(t,f) = \int_{-\infty}^{+\infty} x(\tau)\, w(\tau - t)\, exp(-j2\pi f\tau) d\tau \tag{1.18}$$

where $w(\tau - t)$ is the window function centered at time t and the segmented signal (or observed segment through this window) is $x(\tau)\, w(\tau - t)$.

The resolution trade-off between time and frequency (Due to the uncertainty principle), which is determined by the width of the window is the major disadvantage of the STFT. Although a large width of the window creates good resolution in the frequency domain, it provides poor resolution in the time one. Thus, it is can analyse the signals are stationary at the scale of the short time window. However, it is possible to choose the window in such a way as to minimise the negative consequences of the trade-off; the optimum performance is obtained for the Gaussian window and this leads to the Gabor transform [55].

The energy density spectrum ($|STFT_x(t,f)|^2$) explains how the energy of a signal has been distributed with frequency. The ensemble of the spectra obtained for different time creates the time-frequency distribution called spectrogram. Since the spectrogram has resolution limitations as well, it can only be applied to quasi-stationary signals.

Spectral kurtosis

Based on the STFT, Spectral Kurtosis (SK) provides a means of determining which frequency bands contain a signal of maximum impulsivity and gives a measure of the impulsiveness of a signal as a function of frequency. The SK was first applied to bearing faults in [56,57]. This may be simply detected by computing the kurtosis of the complex envelope $X(t,f)$ as:

$$K(f) = \frac{\langle |X(t,f)|^4 \rangle}{\langle |X(t,f)|^2 \rangle^4} - 2 \tag{1.19}$$

with $\langle \cdot \rangle$ the time-averaging operator and where the subtraction of 2 is used to enforce $K(f) = 0$ in the case $X(t,f)$ is complex Gaussian (instead of 3 for real signals) [56].

The spectral kurtosis is illustrated in Fig. 1.7 for a rolling element bearing signal $x(t)$ modelled as a series of impulse responses.

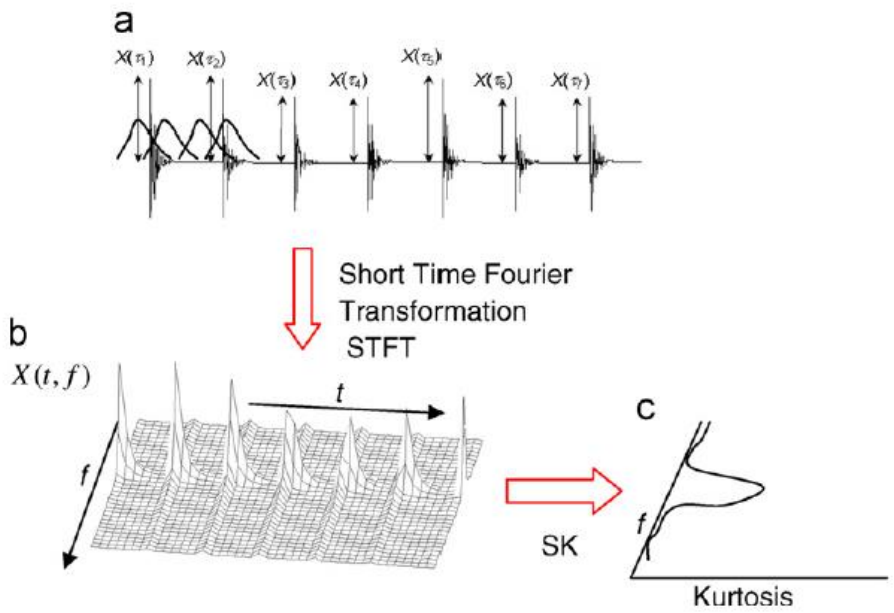


Figure 1.7- Calculation of SK from STFT for a simulated bearing fault signal: (a) simulated time signal, (b) STFT and (c) SK as a function of frequency (from [30]).

Wavelet transform (WT)

The wavelet transform applies wavelets (as the basis functions), instead of sinusoidal functions, as the basis. It adds a scale variable in addition to the time variable in the inner product transform. The continuous wavelet transform (CWT) is usually defined as follows:

$$W_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\, \psi^{*}\left(\frac{t-b}{a}\right) dt \tag{1.20}$$

where b is a translation parameter (is the time shift), a is a dilation or scale parameter and ψ is the wavelet basis (called mother wavelet) dilated and translated.
The signal is decomposed into the base functions created by translation and dilation from the mother wavelets $\psi(t)$. There are a number of different functions utilized as mother wavelets given in [58-60]. For higher frequency components, wavelet transform has a better time localization but a lower frequency resolution. On the other hand, for lower frequency components, the frequency resolution is higher whereas the time localization is worse. The Wavelet transform has been widely used in machinery fault diagnosis and signal denoising, which can be found in the reviews [61,62]. However, the question of how to choose a suitable one among them to match the specific signal structure remains an open issue.

Various types of the Wavelet transform such as Continuous wavelet, Discrete wavelet, Wavelet packet are comprehensively studied in chapter 2.

Order tracking method

In order to diagnose faults in non-stationary working conditions such as time-varying speed, the order tracking method is applied to transform a measured signal from time domain, which is

non-stationary) to angular (or order) domain, which will be stationary. Thus, the smearing problem of discrete frequency components due to speed fluctuations can be avoided. Three main families of order tracking techniques are the following: resampling methods (computed order tracking method), the Kalman filter based methods and the transformed based methods. The Vold-Kalman filter which could estimate amplitude of a series of harmonics of the shaft speed and instantaneous speed is a specific form of Kalman filter. The transform based method performs order tracking and the FT so that the amplitude and phase of each order of the spectrum are estimated.

In the computed order tracking method (COT) technique both the vibration signal and the tachometer pulse are acquired at constant time intervals using conventional hardware. The signal sampled at constant time intervals ($x[n\Delta t]$) is resampled to obtain new data that are sampled at constant angular increments of the shaft's rotation ($\Delta\theta$). These new data are then processed using traditional FFT analysis. In this way, the frequency domain is changed to the order domain [54]. The Fourier transform of the signal from the angular domain is obtained as follows:

$$X(\Omega) = \int_{-\infty}^{+\infty} x(\theta)\, e^{-j\Omega\theta}\, d\theta \tag{1.21}$$

$$X[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n\Delta\theta] e^{-j\Omega[k]n\Delta\theta} \tag{1.22}$$

where $\Delta\theta$ angular domain resolution, N is the number of samples of the interpolated signal $x(\theta)$ and $\Omega[k]$ indicates the vector of orders for the representation of the order spectrum.

The order tracking method has been widely used for fault detection when there exist speed variations. However, there still exist two drawbacks: limitation of speed variation and measurement cost. Many studies have been carried out to improve it [64-69].

Wigner-Ville distribution (WVD)

The approach of the Wigner distribution [70,71] is based on the use of the autocorrelation function for calculating the power spectrum and is defined as:

$$WVD_x(t,f) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)\, x^*\left(t-\frac{\tau}{2}\right)\, exp(-j2\pi f\tau)\, d\tau \tag{1.23}$$

As the signal exists twice in Eq. 1.23, the distribution is called bilinear distribution. Although the WVD has the highest time-frequency resolution, it is a non-linear distribution, which has undesirable impact on the frequency representation of real multi-component signals. Fault diagnosis applications of the Wigner–Ville distribution can be found in [72-75].

Hilbert-Huang transform (HHT)

As the instantaneous frequency of nonstationary multi-components signals usually is a time-variant feature, the accurate estimation of instantaneous frequency is extremely important. Whereas, for mono-tonic signals, the instantaneous frequency can be obtained utilizing the derivative of phase relative to time, the same does not hold for multi component. Therefore, first, the signal is decomposed into several intrinsic mode functions (IMFs) using empirical mode

decomposition (EMD) [76]. Then, by applying the Hilbert transform, the instantaneous frequency and amplitude of each IMFs are estimated.
The decomposed multi component signal $x(t)$ can be represented as:

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{1.24}$$

where the residue r_n represents the mean trend of the signal. In the whole IMFs (c_j), which are zero-mean oscillatory function, the number of extrema and the number of zero crossings must either equal or differ at most by one.
Having obtained the IMFs using EMD method, the Hilbert transform is applied to each IMFs:

$$H[c_i(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{c_i(t)}{t-\tau} d\tau \tag{1.25}$$

Then, analytical signal is constructed as:

$$z_i(t) = c_i(t) + jH[c_i(t)] \tag{1.26}$$

or

$$z_i(t) = a_i(t)\, exp\big(-j2\pi f_i(t)\big) \tag{1.27}$$

Finally, the Hilbert-Huang transform and the Hilbert energy spectrum are represented, respectively as follows:

$$TFR_x(t,f) = Re\left\{\sum_{i=1}^{n} a_i(t)\, exp\left(j\int 2\pi f_i(t)dt\right)\right\} \tag{1.28}$$

$$TFR_x(t,f) = \sum_{i=1}^{n} {a_i}^2(t)\delta[f - f_i(t)] \tag{1.29}$$

where $\delta(\cdot)$ is the Dirac delta function.

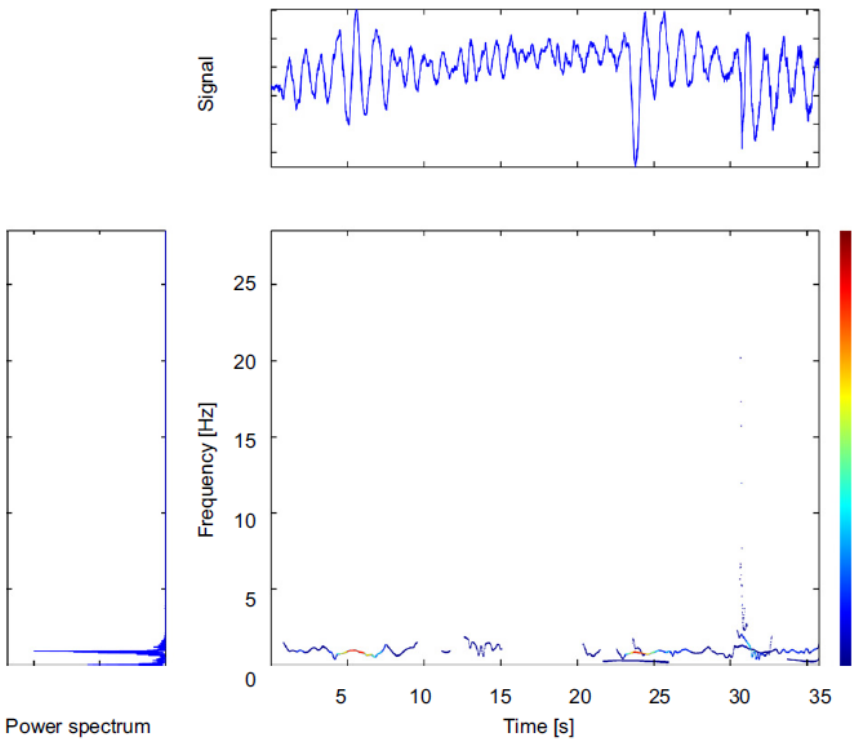


Figure 1.8- Hilbert energy spectrum (from [53])

The HHT of the draft tube hydraulic fluctuation signal of a hydro-turbine during a start-up process is illustrated in Fig. 1.8.

Although the HHT has some advantages, a higher sampling frequency is required (because of using interpolation in the EMD). Furthermore, as the mode mixing problem is a deficiency of the EMD, the ensemble empirical mode decomposition (EEMD), which will be discussed in chapter 2, needs to apply to resolve the problem.

The Huang transform has been widely applied to fault diagnosis of machinery [77-93].

Teager-Huang transform (THT)

The Teager-Kaiser energy operator (TKEO), which is a nonlinear differential operator can estimate the energy required to generate a signal [94-97].

The TKEO is defined for a continuous time signal $x(t)$ as:

$$\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\,\ddot{x}(t) \tag{1.30}$$

where $\dot{x}(t)$ and $\ddot{x}(t)$ are the first and the second time derivatives of $x(t)$, respectively.

For a discrete time signal $x(n)$ (where n is the discrete time index), using difference to approximate differential, the TKEO can be proposed as:

$$\Psi[x(n)] = x^2(n) - x(n+1)\,x(n-1) \tag{1.31}$$

As at any instant, only three consecutive samples are needed to estimate the instantaneous TKEO, it is adaptive to the instantaneous changes in signals and is quite adapt to resolve transient events. It is an adaptive method and effective in estimating the instantaneous frequency and envelope amplitude of non-stationary signals. It has some merits such as low computational cost, high resolution of time and frequency and adaptability to instantaneous feature [53].

The instantaneous frequency and instantaneous amplitude at any time instant of the signal $x(n)$ can be given as:

$$f(n) = \frac{1}{2}\, arccos\left(1 - \frac{\Psi[x(n+1) - x(n-1)]}{2\Psi[x(n)]}\right) \tag{1.32}$$

$$|a(n)| = \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1) - x(n-1)]}} \tag{1.33}$$

In this method, the signal is decomposed into some IMFs by the EMD algorithm. Then, the envelope amplitude and instantaneous frequency of the IMFs are obtained utilizing the TKEO method. The time-frequency distribution of the rotor vibration signal of a real industrial hydro-turbine during a shut-down transient process is shown in Fig. 1.9.

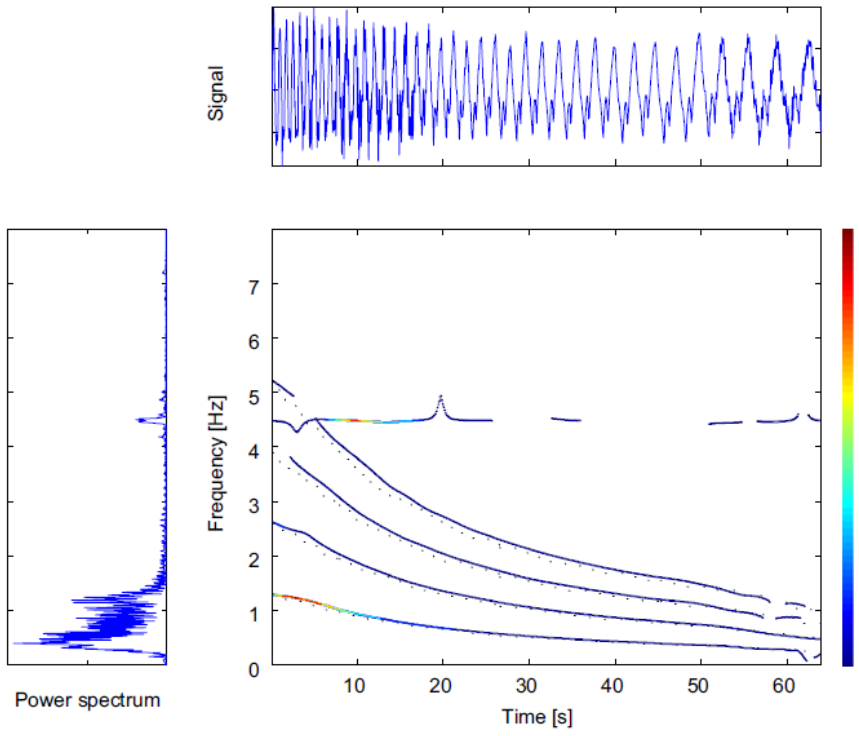


Figure 1.9- Time-frequency Teager energy distribution (from [53]).

The Teager-Kaiser energy operator have been applied in various machinery fault diagnosis [98-106].

Feature selection procedures

The resulting low-dimensional data set from feature extraction is the feature vector or pattern vector used by the pattern recognition algorithm to assign a class. Feature vectors should also be of the lowest dimension possible, as the learning problem becomes much more difficult for high-dimensional input spaces; in particular, much more training data is needed. The feature extraction does not necessarily lead to reduced dimensionality. There exist many feature selection procedures either employ feature reduction techniques and combining existing features into new features or select subset of features [29]. This is the process of selecting those components of a feature vector which carry all (or most of) the discriminatory power of the feature. The remainder are discarded and the dimension of the vector is thereby sometimes reduced- considerably.

Then, data compression methods can be applied to selected features in order to reduce the dimension of data. It can be implemented through transformation such as the Fourier transform or the wavelets or methods like principal component analysis (PCA) [29]. Furthermore, the principal component analysis (PCA) can be used as a decomposition method that decomposes a signal from low frequency bands to high frequency bands. It transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. It provides a linear transform (Eq. 1.34) of a certain number n of m-dimensional input vectors $\boldsymbol{x}_i$ into new vectors $\boldsymbol{s}_i$.

$$\boldsymbol{s}_i = \mathbf{V}^T \boldsymbol{x}_i \qquad i = 1, \dots, n \qquad m < n \tag{1.34}$$

where $\mathbf{V}$ is a m × m orthogonal matrix where on each column $\mathbf{v}_\mathrm{j}$ is present the eigenvector of the sample covariance matrix.

$$\mathrm{C} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \, \boldsymbol{x}_i^T \tag{1.35}$$

Thus the PCA solves the eigenvalue problem:

$$\lambda_j \mathbf{v}_j = \mathrm{C}\, \mathbf{v}_\mathrm{j} \qquad j = 1, \dots, m \tag{1.36}$$

where λ_j denoting the eigenvalues of C and $\mathbf{v}_\mathrm{j}$ the corresponding eigenvector (the eigenvector of the sample covariance matrix.

When the orthogonal transformation of the input vector $\boldsymbol{x}_i$ is computed, the various components of $\boldsymbol{s}_i$ (principal components) are obtained:

$$\boldsymbol{s}_i(j) = \mathbf{v}_j^T \boldsymbol{x}_i \quad j = 1, \dots, m \tag{1.37}$$

If the eigenvalues are sorted in descending order, it is possible to select only the first among them in order to reduce the number of principal components in $\boldsymbol{s}_i$.

1.3.4 Post-processing

In this step the feature vectors are prepared for the pattern recognition stage. It may be only selection of a normalisation for the features or implementing more advanced processing; for example the feature data may be nonlinearly transformed in order to produce data with a Gaussian probability density function [29].

1.3.5 Pattern recognition

The next stage is pattern recognition which is the stage that a method is applied to decide the damage state based on the feature vectors extracted by signal processing techniques. If reliable intelligent monitoring systems could be constructed, then the system monitored can operate at the margin of safety without extended periods of inspection.

The intelligent damage identification system consists of four hierarchical levels [1]:

1. **Detection**: the method gives a qualitative indication that damage might be present in the structure.
2. **Localisation**: the method gives information about the probable position of the damage.
3. **Classification**: the method gives information about the type of damage.
4. **Assessment**: the method gives an estimate of the extent of the damage.
5. **Prediction**: the method offers information about the safety of the structure for example, estimates a residual life.

The level 5 (Prediction) requires a priori knowledge of the physics of the damage, i.e. characterisation. The detection level (Level 1) can be carried out with no prior knowledge of the behaviour of the system when it is damaged.

Many modern approaches to damage identification are based on the idea of pattern recognition. There exist two types of learning algorithm, supervised learning, in which the model is trained by introducing the desired label for each data set. In some cases, due to lack of data corresponding to damage condition, it is not easy to apply supervised learning techniques. Thus, unsupervised algorithm that training data only from the healthy condition of the system is used.

Depending on the desired diagnosis, there exists two types of algorithms:

1. **Novelty detection**. In this case, the algorithm recognizes whether the data comes from normal operating condition or not. The advantage is that unsupervised learning can be used. It refers to lots of practical situation in which class labels are not available. Methods for novelty detection include: outlier analysis [107], kernel density methods [108], autoassociative neural networks [109], Kohonen networks [110], growing radial basis function networks [111], methods based on SPC control charts [112] and one-class support vector machine [113].
2. **Classification**. In this case, the classifier is trained using a supervised learning method to construct a predictive model for normal and faulty classes. There exist some classification methods such as Bayesian classification methods, nearest-neighbour search, artificial neural network classifiers [114, 115] and recently support vector machines [116].

There is a trade-off between the resolution of the diagnosis and the noise rejection capabilities of the algorithm. As the clean data has little fluctuation in the measurement from normal operating condition, small damages will cause detectable deviations. Eliminating fluctuations on the healthy condition data, as far as possible, is necessary for intelligent fault detection.

1.4 Model-based approaches

Another damage identification approach is the model-based approaches [117-119]. These approaches create explicit mathematical model of the system, which can be effective if a correct and accurate model is built. However, it would be very difficult or even impossible to build mathematical models for complex systems.

The two most common model based approaches are the residual view and the parameter estimation.

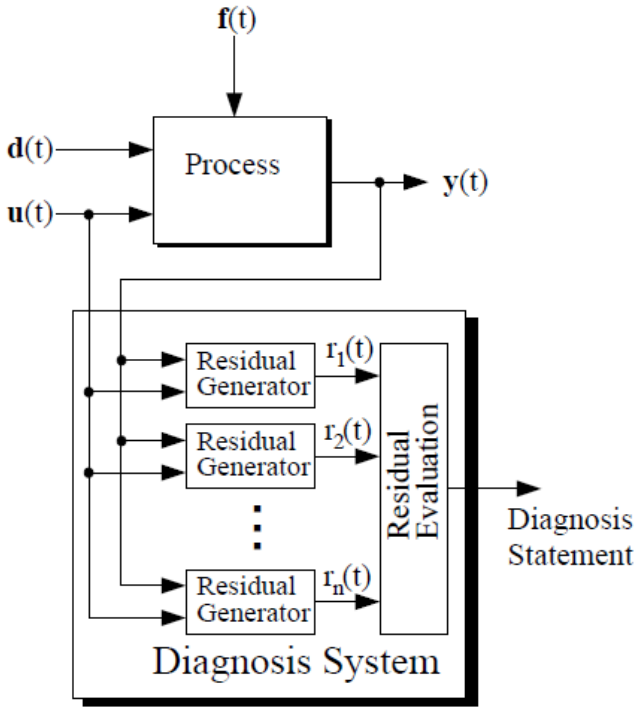


Figure 1.10- A diagnosis system based on the residual view (from [119]).

The Residual view

In this approach illustrated in Fig. 1.10, $f(t)$ is the model of the faults and the residual $r(t)$, which is a scalar or vector is zero or small value in the healthy condition($f(t) = 0$), and is not zero when a fault occurs $f(t) \neq 0$. The diagnosis system is then separated into two parts: residual generation and residual evaluation.

Parameter estimation

In the Parameter estimation approach faults are modelled as deviations in constant parameters. In this approach, an estimated parameter, obtained using parameter estimation techniques, is compared with the nominal value. If the estimated parameter deviates too much from the nominal value, then the conclusion is that a fault has occurred.

The disadvantage of model based diagnosis is quite naturally the need for a reliable model and possibly a more complex design procedure. Various model-based diagnostic approaches have been

applied to fault diagnosis of mechanical systems such as gearboxes [120-122], bearings [123–126], rotors [127,128].

Chapter 2

Diagnostic techniques

In this chapter pre-processing and processing techniques applied to the signals for damage detection are presented. As mentioned in section 1.3.2, the collected data would be polluted by different sources of noise. In order to prepare the data for feature extraction, denoising can be applied to remove noise. There exist different approaches for denoising, which are discussed in section 2.1. After cleaning the data, features are extracted using different processing methods. Decomposition a complicated signal into some elementary components is a common method to extract features. Empirical mode decomposition (EMD), Ensemble empirical decomposition (EEMD), Teager-Kaiser energy operator (TKEO) and cointegration are the methods introduced in section 2.2.

2.1 Denoising

The collected data could be polluted by the sources such as the measurement system and other sources unrelated to the investigated system. Data cleansing is required to treat the data for feature extraction. There exist some approaches for noise cancelling such as classical approach (e.g. Kalman filtering), adaptive filtering and Wavelet transform approaches. The Kalman filter is an optimal estimator used for linear systems. Adaptive filtering try to separate two uncorrelated components (e.g. deterministic and random components) using a reference signal. In the wavelet transform approach, the signal is decomposed through highpass and lowpass filters (in the case of wavelet packet decomposition). Then, the predetermined threshold is applied so that the impulses can be retained effectively.

2.1.1 Kalman filter

This filter is named from Kalman, one of the primary developers of this theory [129]. The Kalman filter performs a predict-correct estimator so that minimizes the estimated error covariance. In the predict step, the state estimate from the previous time step (at time $k-1$) is used to the estimate the state at the current time step (at time k).

The discrete state equation of a time-variant system, which allows to predict the future, is formulated as:

$$x_k = A_{k-1}\, x_{k-1} + B_{k-1}u_{k-1} + v_{k-1} \qquad k = 0,1, \ldots \tag{2.1}$$

With the output measurement z_k as follows:

$$z_k = C_{k-1}x_{k-1} + w_{k-1} \tag{2.2}$$

where $x\epsilon\mathrm{R}^n$ is the state vector, $u\epsilon\mathrm{R}^m$ is the known input vector, A is the $n \times n$ state transition matrix, $v\epsilon\mathrm{R}^n$ is unknown process noise ($v \sim N(0, Q)$). $z\epsilon\mathrm{R}^m$ is the measurement, C indicates the $m \times n$ measurement matrix and $w\epsilon\mathrm{R}^m$ is the measurement noise ($v \sim N(0, R)$).

The filtering equations can be represented in two categories: time update equations and measurement update equations, which can be remarked as predict and correct equations as well. In the time update equations, the current state and error covariance estimates (at time k-1) are utilized to obtain the a priori estimates (at time k).

The time update equations are as follows [130]:

$$\hat{x}_{k|k-1} = A_k \, \hat{x}_{k-1|k-1} + B_k \, u_k \tag{2.3}$$

$$P_{k|k-1} = A_k \, P_{k-1|k-1} \, A_k^T + D_k \tag{2.4}$$

where $\hat{x}_{k|k-1}$ represents the state estimate at time k based on the measurements up to and including at time k, $P_{k|k-1}$ is the error covariance matrix (a measure of the estimated accuracy of the state estimate) and D indicates the covariance of the process noise.

In the measurement update the a priori state estimate and the new measurement (at time k) is combined to obtain an improved a posteriori estimate. The equations are as follows:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \, (z_k - C_k \, \hat{x}_{k|k-1}) \tag{2.5}$$

$$P_{k|k} = (I - K_k \, C_k) \, P_{k|k-1} \tag{2.6}$$

where $\hat{x}_{k|k}$ indicates the state estimate at time k based on the measurements up to and including at time k, $P_{k|k}$ denoting the error covariance matrix (a measure of the estimated accuracy of the state estimate) and K is the optimal gain of the standard Kalman filter defined as follows:

$$K_k = P_{k|k-1} \, C_k^T \, (C_k \, P_{k|k-1} \, C_k^T + R_k)^{-1} \tag{2.7}$$

where R is the covariance of the observation noise.

In the measurement update, first the Kalman gain is calculated (Eq. 2.7), then an *a posteriori* state estimate and an *a posteriori* error covariance estimate are computed. After each time and measurement update pair, the recursive procedure is repeated with the previous *a posteriori* estimates used to predict the new a priori estimates.

Wu et al. [131] applied the order tracking fault diagnosis technique based on the Kalman filtering algorithm to gear-set defect diagnosis. Li et al. [132] performed the Kalman filter for denoising the vibration signals of rolling bearings. Khanam et al. [133] mixed some simulated noise and external vibrations with a clean experimental signal to investigate efficiency of the Kalman filter under various noisy environments.

2.1.2 Adaptive noise cancellation

In rotating machines, especially in gearbox (even in healthy gears), there exist strong discrete frequency noise to contaminate frequency bands where the bearing signal could dominant [30]. In order to remove such discrete frequency noise, some methods such as adaptive noise cancellation

(ANC) and self- adaptive noise cancellation (SANC) can be utilized. An adaptive filter is a time-variant filter, which modifies the transfer function (filter coefficients) to optimize a predefined criterion (cost function) [134]. Two uncorrelated components can be separated using a reference signal as shown in Fig. 2.1. The modified reference signal is subtracted from the original signal to remove the uncorrelated component [30].

The cost function may defined as the mean squared estimation error:

$$CF_k = E\left[\left|d_k - \hat{d}_k\right|^2\right] = E[|\varepsilon_k|^2] \tag{2.8}$$

where $E[\cdot]$ is the expectation, d_k denotes the signal includes the desired component and the uncorrelated component, $\hat{d}_k$ is the estimated or output signal and ε_k is the estimation error.

The coefficients are updated at every time instant k:

$$W_{k+1} = W_k + \Delta W_k \tag{2.9}$$

where W_k is the coefficient of the filter and is represented as $W_k = [w_{0k}, w_{1k}, \dots, w_{Lk}]^T$, where L is the order of the filter.

It can be represented as follows:

$$W_{k+1} = W_k + 2\mu\varepsilon_k W_k \tag{2.10}$$

where μ is the convergence factor.

The output of the filter, the estimated signal, is expressed as:

$$\hat{d}_k = \sum_{l=0}^{L} W_{lk}\, x_k \tag{2.11}$$

where x_k is the reference signal correlated with the desired signal.

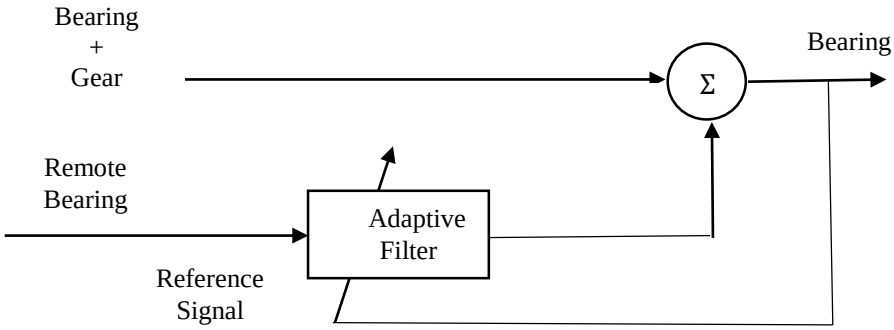


Figure 2.1- Adaptive noise cancellation used for removing the deterministic uncorrelated component (from gear) from desired signal (from bearing)

The ANC was applied to extract a faulty bearing signal measured on a gearbox [135,136]. The reference signal was measured over another remote bearing to be uncontaminated. However, in some cases there is no possibility to acquire an uncontaminated reference signal such as on a planetary gearbox, in which all signals are transmitted through the ring gear [30].

Self-adaptive noise cancellation

When the signal contains a deterministic (discrete frequency) and a random components, self-adaptive noise cancelation (SANC) method can be applied [136]. As illustrated in Fig. 2.2, the separation can be performed using only one signal. The reference signal is built by making a delay on the original signal, longer than the correlation length of the random signal. The relationship between the original signal and the delay is not identified by the filter and a transfer function is found between the delay and the deterministic component of the signal (gear signal).

The least mean squares (LMS) algorithm is expressed as follows [136]:

$$W_{k+1} = W_k - \mu \nabla_k \tag{2.12}$$

where μ is a convergence factor, which is trade-off between avoiding divergence and adaptation time and gradient vector can be formulated as:

$$\nabla_k = \frac{\partial E[\varepsilon_k^2]}{\partial W_k} \tag{2.13}$$

A conservative approximation for Eq. 2.12 is [137]:

$$W_{k+1} = W_k + \frac{2\mu_n \varepsilon_k x_k}{(L+1)\hat{\sigma}_k^2} \tag{2.14}$$

where μ_n is the normalized convergence factor: $0 < \mu_n < 1$, μ is the convergence factor: $\mu = (\mu_n/((L+1)\hat{\sigma}_k^2))$ and $\hat{\sigma}_k^2$ denotes the exponential-averaged estimate of the input signal power at the k-th iteration.

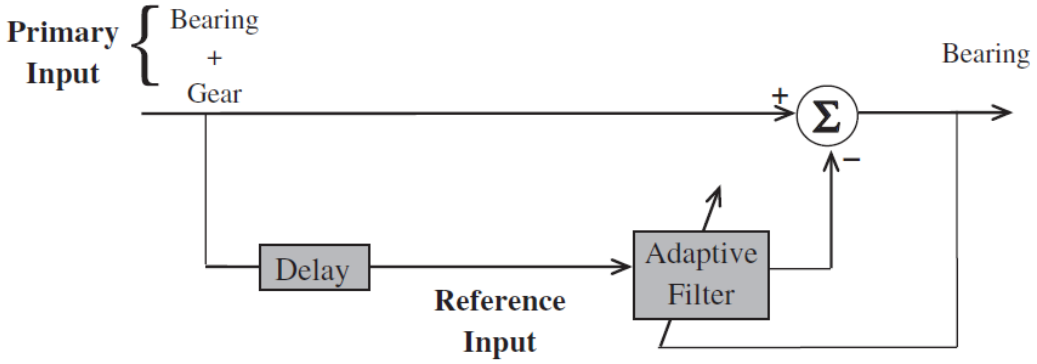


Figure 2.2- Self-adaptive noise cancellation used for removing the deterministic uncorrelated component (from gear) from the desired signal (from bearing) (from [30])

The estimated signal is a linear combination of present and past input samples, which is given by the convolution sum as follows:

$$\hat{d}_k = \sum_{l=0}^{L} W_{lk}\, x_{k-l} \tag{2.15}$$

There exist a few important parameters required to be set up such as filter length L, delay Δ, and convergence factor discussed comprehensively in Ref. [30,138,139]. The most important may be L (the order of the adaptive filter) which, in the case of gear and bearing signals separation, is quite large, typically of a few hundreds and even thousands [30]. The delay should be determined longer than the correlation length of the random component and, on the other hand, should be shorter than that of the deterministic part [30].

Randall and Li applied SANC to extract faulty bearing signal collected on a planetary gearbox [140]. Lee and White [141] utilized the technique to diagnose the faults within an internal combustion engine and also for an industrial gearbox. Barszcz applied it to a real vibration signal from a wind turbine with an outer race fault [142]. Antoni and Randall proposed and optimization of SNAC for separating bearing and gear signals [143].

2.1.3 Wavelet transform (WT)

In this approach the signal is considered as a superposition of components with respect to a set of base functions. These basis functions can then be used to represent different frequency content simply by scaling them with respect to time. Signal decomposition using such scaled functions results in the so called time-scale. In general, this analysis results in a series of wavelet coefficients indicating how close the signal is to the particular wavelet itself.

The WT could be categorized as: the Continuous wavelet transform (CWT), the Discrete wavelet transform (DWT) and the Wavelet packet transform (WPT).

Continuous wavelet transform (CWT)

The continuous wavelet transform (CWT), mentioned in chapter 1, is usually defined as:

$$W_\psi(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\, \psi^* \left(\frac{t-b}{a}\right) dt \tag{2.16}$$

where b is a translation parameter (is the time shift), a is a dilation or scale parameter and ψ is the wavelet basis (called mother wavelet) dilated and translated.

The signal is decomposed into the base functions created by translation and dilation from the mother wavelets $\psi(t)$. There exist a number of different functions utilized as mother wavelets such as Haar, Daubechies (db), Symlets, Coiflets, Gaussian, Morlet, complex Morlet, Mexican hat, bi-orthogonal wavelets, reverse biorthogonal, Meyer, harmonic wavelets, discrete approximation of Meyer, complex Gaussian, Shannon, and frequency B-spline [58-60]; some of them are shown in Fig. 2.3. For higher frequency components, wavelet transform has a better time localization but

a lower frequency resolution. On the other hand, for lower frequency components, the frequency resolution is higher whereas the time localization is worse. The wavelet transform $W_\psi(a,b)$ is normalised by $1/\sqrt{a}$ to ensure that the integrated energy is independent of the dilation.

Discrete wavelet transform (DWT)

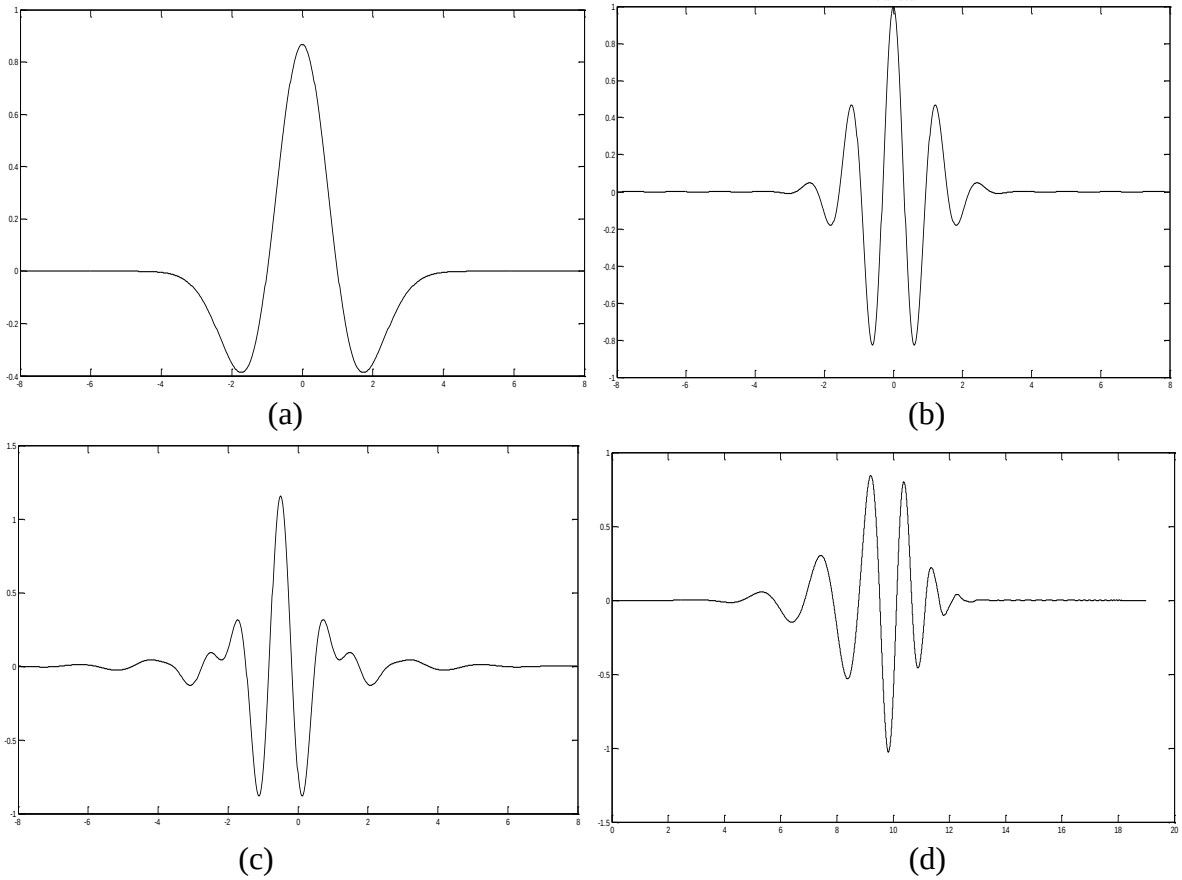
Applying the CWT to a signal will generate redundant information since the scale and translation parameter are changed continuously. Supposing that the number of time points is $N = 2^J$, it can be shown that the signal can be completely reconstructed from discrete samples on the dyadic time-scale given by the scale a and the time b are discretized as following [29]:

$$a = 2^{-j}, b = k\,2^{-j} \tag{2.17}$$

where j is integer from 1 to $J-1$ and k is integer in the range1 to 2^j. The higher values of j correspond to shorter time scales and thus to higher frequencies.

The discrete wavelet transform (DWT) can be defined as:

$$w_k^j = \int_{-\infty}^{+\infty} x(t)\psi_{j,k}^*(t)\,dt \tag{2.18}$$



(a) (b) (c) (d)

Figure 2.3- Some mother wavelets: a) Mexican hat, b) Morlet, c) Meyer and d) Daubuchi 10

where $\psi_{j,k}(t)$ are translated and dilated mother wavelet ψ defined in terms of the dyadic grid as follows:

$$\psi_{j,k}(t) = 2^{j/2}\,\psi\left(2^{j}t - k\right) \tag{2.19}$$

The original signal can be reconstructed as follows:

$$x(t) = \sum_{j=0}^{J-1}\sum_{k=1}^{2^{j}} w_k^{j}\,\psi_{j,k}(t) \tag{2.20}$$

The DWT process can be considered as applying a low-pass and high-pass filter (known as Quadrature mirror filters (QMF)), denoted as $h(k)$ and $g(k) = (-1)^{k}h(1-k)$, respectively. The detail is considered as the result of the high-pass and the approximation as the result of the low-pass filter.

These filters are expressed as follows [144]:

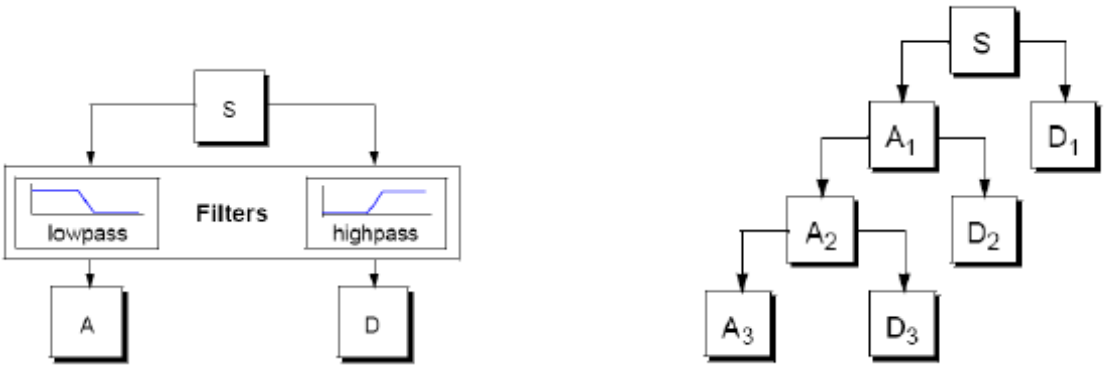
$$\begin{cases} \phi(t) = \sqrt{2}\sum_{k} h(k)\phi(2t-k) \\ \psi(t) = \sqrt{2}\sum_{k} g(k)\phi(2t-k) \end{cases} \tag{2.21}$$

with $\sum_k h(k) = \sqrt{2}$ and $\sum_k g(k) = 0$.

Applying the DWT, the signal is decomposed into a set of low and high- frequency components as [144]:

$$\begin{cases} a_{j,k} = \sum_{m} h(2k-m)\,a_{j-1,m} \\ d_{j,k} = \sum_{m} g(2k-m)\,a_{j-1,m} \end{cases} \tag{2.22}$$

The decomposition procedure is illustrated in Fig. 2.4.



(a) (b)

Figure 2.4-. Filtering process: (a) the signal (S) is decomposed into approximation (A) and detail (D) components; (b) wavelet decomposition tree (from [17]).

Wavelet packet transform (WPT)

A signal is decomposed into approximation and detail through low and high band filters, respectively. Then both of them are decomposed into a second level approximation and detail and the process is repeated as shown in Fig. 2.5.

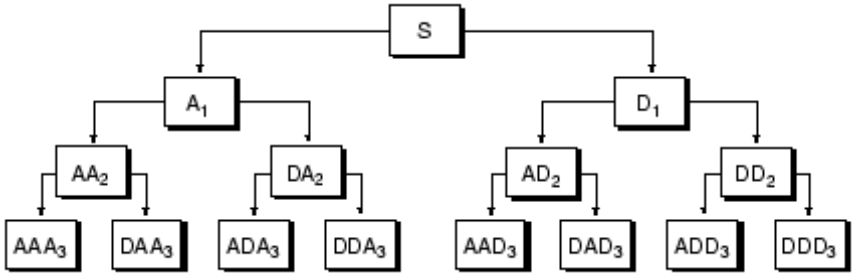


Figure 2.5- Wavelet packet decomposition tree (from [17]).

The basic wavelet packet functions are defined as [16]:

$$\begin{cases} u_{2n}(t) = \sqrt{2}\sum_k h(k)u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2}\sum_k g(k)u_n(2t-k) \end{cases} \tag{2.23}$$

where $u_0(t) = \phi(t)$ and $u_1(t) = \psi(t)$.

The WPT decomposes the signal as follows:

$$\begin{cases} d_{j+1,2n} = \sum_{m} h(m-2k)\, d_{j,n} \\ d_{j+1,2n+1} = \sum_{m} g(m-2k)\, d_{j,n} \end{cases} \tag{2.24}$$

where $d_{j,n}$ denotes the wavelet coefficients at the j level, n sub-band, $d_{j+1,2n}$ and $d_{j+1,2n+1}$ are the wavelet coefficients at the level of $j+1$, $2n$ and $2n+1$ sub-bands, respectively, and m is the number of the wavelet coefficients.

In general, the CWT is more suitable for time-frequency analysis and visualization, and the DWT is more appropriate for decomposition, compression and feature selection. The WPT is faster than the CWT technique, as it uses orthogonal and bi-orthogonal bases with a better resolution in the high-frequency region [146].

In the WT approaches, the selection of the mother wavelet function plays a big role for its efficiency. There exist some research to investigate the most suitable mother wavelet in different applications. Rafiee et al. [147-149], in a thorough study of mother wavelet choice investigation, studied 324 candidate mother wavelet functions from various families including Haar, Daubechies (db), Symlet, Coiflet, Gaussian, Morlet, complex Morlet, Mexican hat, bio-orthogonal, reverse bio-orthogonal, Meyer, discrete approximation of Meyer, complex Gaussian, Shannon, and frequency Bspline.

From a simulation point of view, it is found that Daubechies and discrete Meyer are the best wavelets to be used in vibration signal analysis [146].

Denoising by Wavelet transform

In general filtering methods, the frequency components beyond a certain limit are often set to zero. This may cause some useful fault information to be lost, since faults often appear as impulses which often cover a wide frequency range. Instead, wavelet transform (WT) assigns big wavelet coefficients to such impulses and then, by setting small coefficients to zero, impulses can be effectively retained [61]. The DWT decomposes just the previous approximation coefficients through high and low pass filters, whereas WPT decomposes both the detail and approximation coefficients, which keeps the important information available in higher frequency components.

The denoising method by WPT is represented in the following steps:

a) Decomposition:

A signal is decomposed into approximation and detail using Eq. 2.24. Since the number of decompositions from a signal may be very large, it is necessary to find a method to decide which components must be taken to have an accurate enough representation of the signal. There exist various entropy types for calculating the lowest cost basis, such as Shannon, log energy, sure. One of the most frequently encountered functions is Shannon entropy, which is a measure of signal complexity to wavelet coefficients. Higher entropy values indicate higher uncertainty and consequently higher complexity. If the entropy value is greater than one, and splitting makes the entropy decrease, the component has the potential to reveal more information about the signal and it needs to be decomposed further in order to obtain simple frequency components of the signal [150]:

$$E(s_i) = -{s_i}^2 \log({s_i}^2)$$

(2.25)

$$E(s) = -\sum_i {s_i}^2 \log({s_i}^2)$$

(2.26)

where s is the signal and s_i are the coefficients of the wavelet decomposition of s.

b) Processing the coefficient using thresholding:

Wavelet thresholding is based on the idea that the energy of the signal is concentrated in a few wavelet coefficients, while the energy of noise spreads throughout all the wavelet coefficients. Thresholding is aimed at removing the components caused by noise and is divided into two steps. Firstly, an algorithm is defined to calculate a threshold and then it is necessary to determine how to apply the threshold to the data and remove the irrelevant parts in the wavelet coefficient. There are some algorithms proposed to determine the threshold, such as Universal threshold [151], SURE threshold [152] and the Minimax threshold [151]. The Universal threshold rule is used to process the simulation of the signal series in the following form [151]:

$$T = \sigma\sqrt{2\ \log N}$$

(2.27)

where N is the signal length and σ is the noise standard deviation. The true value of the noise standard deviation σ is, generally, unknown. It is often estimated by $\sigma = MAD/0.6745$, where MAD denotes the median absolute value of the finest scale wavelet coefficients.

The second step is probably the most critical and has quite an impact upon the effectiveness of the procedure, in which thresholding is applied to data. It could be hard thresholding or soft thresholding [153]:

$$y_g = \begin{cases} sgn(y)(|y| - \alpha T) & |y| > T \\ 0 & |y| < T \end{cases}$$

(2.28)

where T is the threshold and α is a constant, which is $0 \leq \alpha \leq 1$. For $\alpha = 0$, it becomes Hard-thresholding, and when $\alpha = 1$, it becomes Soft-thresholding, as illustrated in Fig. 2.6. For cases where precise recovery of signal magnitude is not required, for example, image noise reduction, the soft thresholding is widely used since it can retain the regularity of signal [154].

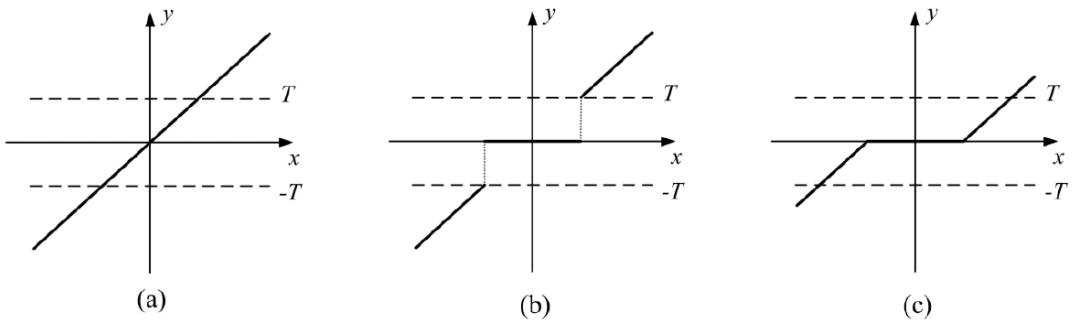


Figure 2.6- a) the signal, b) Hard thresolding and c) Soft thresholding (from [153])

The combination of the soft thresholding policy and universal threshold is referred to as VisuShrink. It ensures a noise-free reconstruction but often the threshold is set too high. Donoho and Jonestone [151] introduced the "minimax" threshold to enhance the universal threshold. The "minimax" threshold level can be much lower than the universal threshold level when it comes to small-to-moderate sample sizes.

c) Reconstruction:

The denoised signal can be calculated iteratively:

$$x = \sum_{k} h(m-2k)\, d_{j+1,2n} + \sum_{k} g(m-2k)\, d_{j+1,2n+1} \tag{2.29}$$

where $d_{j+1,2n}$ and $d_{j+1,2n+1}$ are the wavelet coefficients at the level of $j+1$, $2n$ and $2n+1$ sub-bands, respectively, and m is the number of the wavelet coefficients.

Application

The merits of WT have led it to be used widely in fault diagnosis of rotating machines, which were presented in two comprehensive reviews [155,156]. The CWT and DWT based denoising has been widely adopted in many studies to remove noise in the signals and extract fault-related information.

The first foundations in wavelet-based denoising were implemented by Donoho [154]. Qiu et al. optimized the Morlet wavelet shape factor using the Shanon entropy. They performed singular value decomposition (SVD) to detect periodicity of the signal (from faulty bearings) for appropriate decomposition scale selection [157]. Abbasion et al. [158] studied the condition of an electric motor with two rolling bearings and utilized the CWT (Meyer wavelet) for denosing the collected signal. Jafarizadeh et al. combined the time-averaging-based noise cancellation with Morlet wavelet filtering to identify gear damage [159]. Al-Raheem et al. optimized the CWT for denoising the signals measured from faulty bearings to detect the faults [160]. He et al. applied the Morlet wavelet filtering with sparse code shrinkage to enhance the impulsive features and suppress residual noise for bearing fault identification [161]. Su et al. [162] optimized the parameters of the Morlet wavelet using genetic algorithm; it was shown that with enhancement of autocorrelation to the wavelet coefficients, it is very effective for bearing faults detection. Chiementin et al. [163] applied the DWT to the signals acquired from faulty bearings and attempted to optimize the various parameters selection. They concluded that the wavelet approach enhanced the signal kurtosis and crest factor more than the other techniques. Li et al. [164] developed a gear multi-fault diagnosis method where the DWT was

carried out to denoise the vibration signals to extract the features by autoregressive (AR) model and principal component analysis (PCA). Altmann and Mathew [165] used the WPT analysis-based multiple band-pass filtering for the vibration signals of a low speed rolling-element bearing and significantly improved results were obtained compared to its high-pass counterpart. Peng et al. [166] used the WPT as a pre-processor to decompose the signal into narrow band signals to improve the performance of Hilbert–Huang procedure.

2.2 Signal processing

After denoising, data are processed to extract features. Decomposition is a common method in signal processing, where a multi component signal is broken down into some elementary components. Some decomposition techniques such as Empirical mode decomposition are introduced in the following sections. In order to extract features, energy operator (e.g. Teager-Kaiser operator) can be applied to the decomposed signal, which is discussed in section 2.2.3. The signals collected under time-varying operating conditions, require the remove non-stationarity created by operational variation conditions. Cointegration, presented in section 2.2.4, extracts stationary linear combination among the non-stationary signals.

2.2.1 Empirical mode decomposition (EMD)

The Empirical mode decomposition method (EMD) proposed in 1998 [76], is a self-adaptive data driven method, which decomposes a complex signal into a number of simple oscillatory modes called intrinsic mode functions (IMFs) (Fig. 2.7).

These IMFs are determined by the signal itself rather than by pre-determined functions and designated by the following definitions:

1) In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one.

2) At any point, the mean value of the envelopes defined by local maxima and the envelope defined by the local minima is zero.

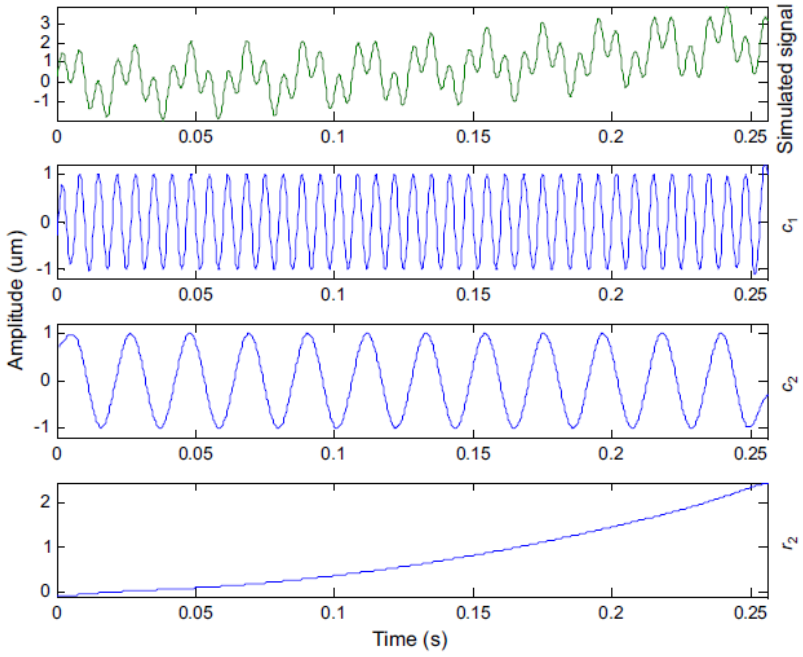


Figure 2.7- A simulated signal and the decomposed IMFs using EMD (from [167]).

The first point means that all maxima must be over the time axes while all minima must be under it. Furthermore, it helps in assuring that the signal is not a wave rising or descending. The second aspect is related, instead, one the signal shape, since it must be symmetric, with time mean locally equal to zero in order to avoid unwanted fluctuations induced by asymmetric waveforms. Each IMF includes different frequency bands ranging from high to low.

The decomposition consists of the following steps:

1) To identify all the local extrema, and then connect all the local maxima by an interpolation method to produce the upper envelope. Repeat the procedure for the local minima to produce the lower envelope.

2) To determine the difference between the signal $x(t)$ and m_1 which is the mean of upper and lower envelopes to obtain the first component, h_1.

$$x(t) - m_1 = h_1 \tag{2.30}$$

If h_1 is an IMF, then it would be the first component of $x(t)$. Otherwise, h_1 is treated as the original signal and step (1)–(2) are repeated:

$$h_1 - m_{11} = h_{11} \tag{2.31}$$

in which, m_{11} is the mean of upper and lower envelope value of h_1. After repeated sifting, up to *k* times (based on sifting stop criterion), h_{1k} becomes an IMF, that is:

$$h_{1(k-1)} - m_{1k} = h_{1k} \tag{2.32}$$

Then it is designated as the first IMF component of the data, $c_1 = h_{1k}$. The sifting process can be stopped by any preselected criterion which will be discussed in the next section.

3) To separate IMF (c_1) from the original signal $x(t)$ to obtain the residue r_1:

$$r_1 = x(t) - c_1 \tag{2.33}$$

4) To consider r_1 as the new data and repeat the above described process for n times, so that n-IMFs of signal $x(t)$ can be obtained. Then:

$$\begin{aligned} r_1 - c_2 &= r_2 \\ &\cdots \\ r_{n-1} - c_n &= r_n \end{aligned} \tag{2.34}$$

5) To stop the decomposition process when r_n becomes a monotonic function from which no more IMF can be extracted. By summing up Equations (4) and (5), we finally obtain:

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{2.35}$$

Thus, a signal $x(t)$ can be decomposed into n zero-mean amplitude modulation frequency waveforms and a residue r_n representing the mean trend of the signal. Going from one residual to the next, the number of extrema is decreased. Thus, the decomposition process is completed with a finite number of modes. A real vibration signal collected from a roller bearing and its decomposed IMFs (c1-c12) are shown in Fig. 2.8.

Flandrin et al. [168] recognized that the EMD acts as a dyadic filter bank (a wavelet-like filter bank) when it was applied to fractional Gaussian noise. This numerical experiment was utilized in order to identify how the EMD behaves in stochastic situations involving broadband noise. A dyadic filter bank is a collection of band-pass filters that have a constant band-pass shape (e.g., a Gaussian distribution) but with neighboring filters covering half or double of the frequency range of any single filter in the bank. The frequency ranges of the filters can be overlapped. Wu and Huang [169] investigated white noise and obtained the similar results.

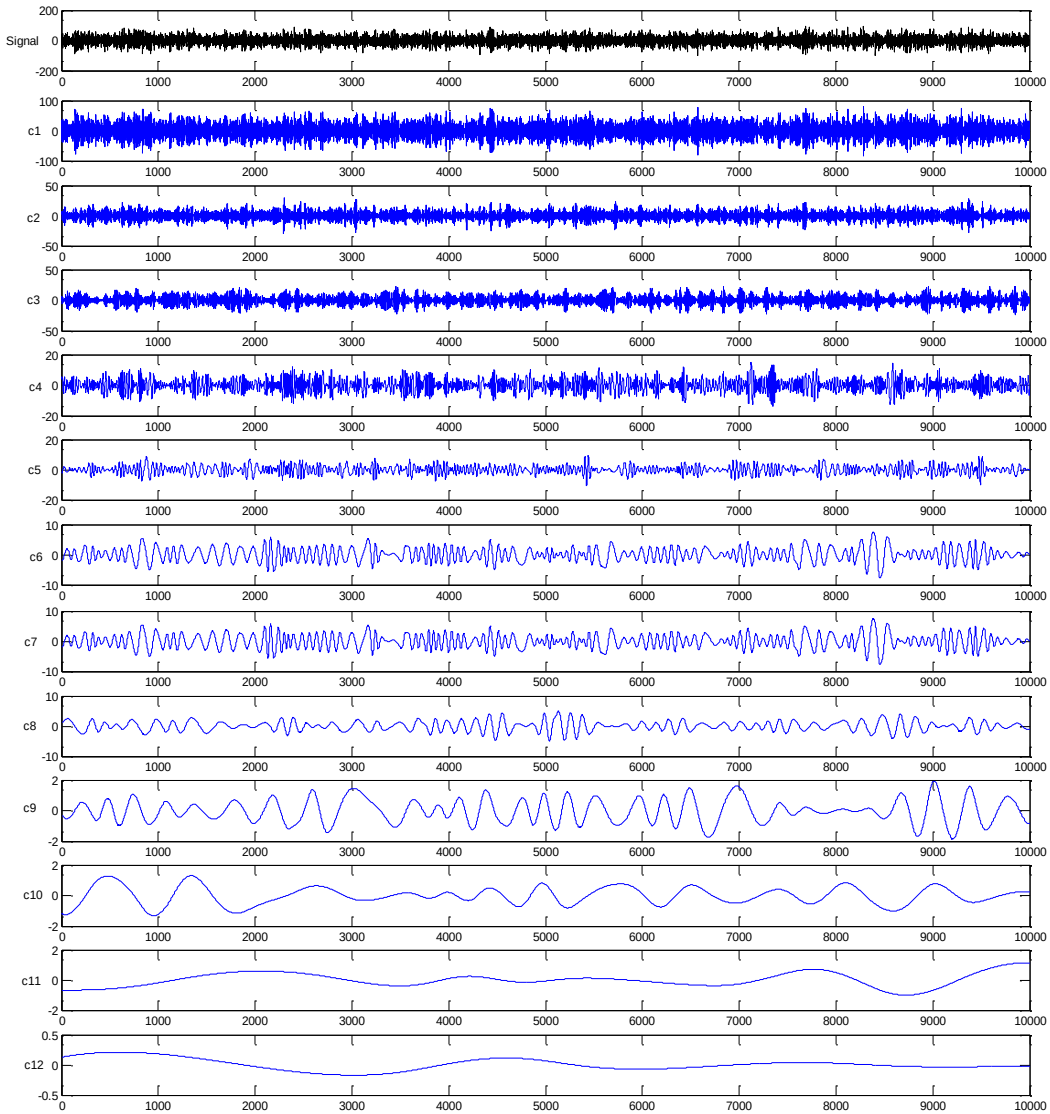


Figure 2.8- A real collected vibration signal of a roller bearing and the decomposed IMFs by EMD

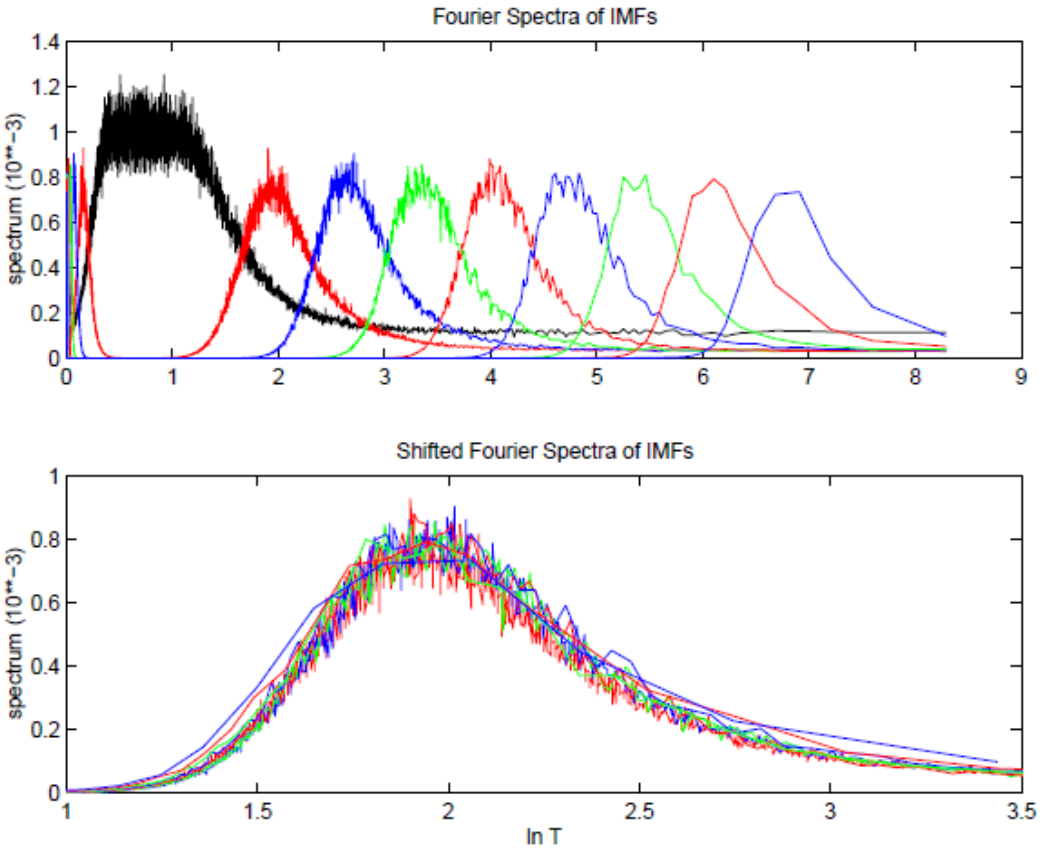


Figure 2.9- The Fourier spectra of IMFs and the spectra as a function of the logarithm of period (from [169])

They found that the EMD acts as a dyadic filter capable of separating the white noise into IMF components having mean periods, each exact twice the value of the previous one. Second, the IMF components are all normally distributed. And third, the Fourier spectra of the IMF components are identical in shape, and cover the same area on semi-logarithmic period scale (as presented in Fig. 2.9). Gao et al. [170] consider the filtering features of EMD and compare it with those of wavelet decomposition. The EMD characteristic of being an adaptive band-pass filter bank with the bandwidth self-adaptively determined by the signal decomposed, whereas the filter features of wavelet decomposition is not self-adaptive.

2.2.2 Ensemble empirical mode decomposition (EEMD)

Although EMD is a powerful tool for signal processing, it has some drawbacks, such as the lack of a mathematical base, no robust stopping criterion for the sifting process, mode mixing and the border effect problem. The intermittency of the detected extrema which belong to the different components is the main reason for mode mixing and make the physical meaning of individual IMF unclear. In order to overcome the problem of mode mixing, Ensemble empirical mode decomposition (EEMD), which is a noise assisted data analysis method was introduced [171]. It is based on the statistical properties of white noise, which showed that the EMD method is an effective self-adaptive dyadic filter bank when applied to the white noise, and the noise could help data analysis in the decomposition of EMD [168,169,172]. The EEMD repeatedly decomposes the original signal with added white noise into a series of IMFs by applying the original EMD process, and treats the means of the corresponding IMFs during the

repetitive process as the final EEMD decomposition result. Since white noise is added throughout the entire signal decomposition process, mode mixing is effectively eliminated. Since the noise in each trial is different, it is cancelled out in the ensemble mean of enough trials.

The idea of adding noise proposed by Flandrin et al.[168] to resolve one of the deficiency of the EMD. If the data lacks the necessary extrema (a Dirac pulse (delta function)), the technique cannot work, as it is based on the existence of extrema. They applied an ensemble of 5000 decompositions using different realizations of noise, as the decomposition results are sensitive to the added noise. Finally, the mean is utilized as the final decomposition of the Dirac pulse.

The data shown in Fig. 2.10 consists of two parts: a low-frequency sinusoidal wave and an impulse high frequency components. The first IMF (decomposed by the EMD) is the mixture of both the low-frequency fundamental and the high-frequency intermittent waves. The EMD could not express the real characteristics of the both components, whereas the two components are accurately decomposed into two IMFs by the EEMD.

The decomposition steps by the EEMD are illustrated in Fig. 2.11 and can be summarized as follows:

a) To add a random white noise signal to the acquired original signal:

$$x_j(t) = x(t) + Amp \cdot n_j(t) \qquad j = 1,2,\dots,M \tag{2.36}$$

where Amp is the amplitude of added white noise and M is the pre-determined number of trial.

b) To decompose the obtained signal ($x_j(t)$) into IMFs using EMD:

$$x_j(t) = \sum_{i=1}^{N_j} c_{ij} + r_{N_j} \tag{2.37}$$

where c_{ij} represents the i-th IMF of the j-th trial, r_{N_j} represents the residue of j-th trial and N_j is the IMFs number of the j-th trial.

c) To repeat steps a and b until the predefined ensemble trial number (M) (add different random noise signal each time).

d) To calculate the ensemble means of the corresponding IMFs of the decompositions as the final result (c_i):

$$c_i(t) = \left(\sum_{j=1}^{M} c_{ij}\right) \Big/ M \qquad i = 1,2,\dots,I \tag{2.38}$$

where I is the minimum number of IMFs among all the trials.

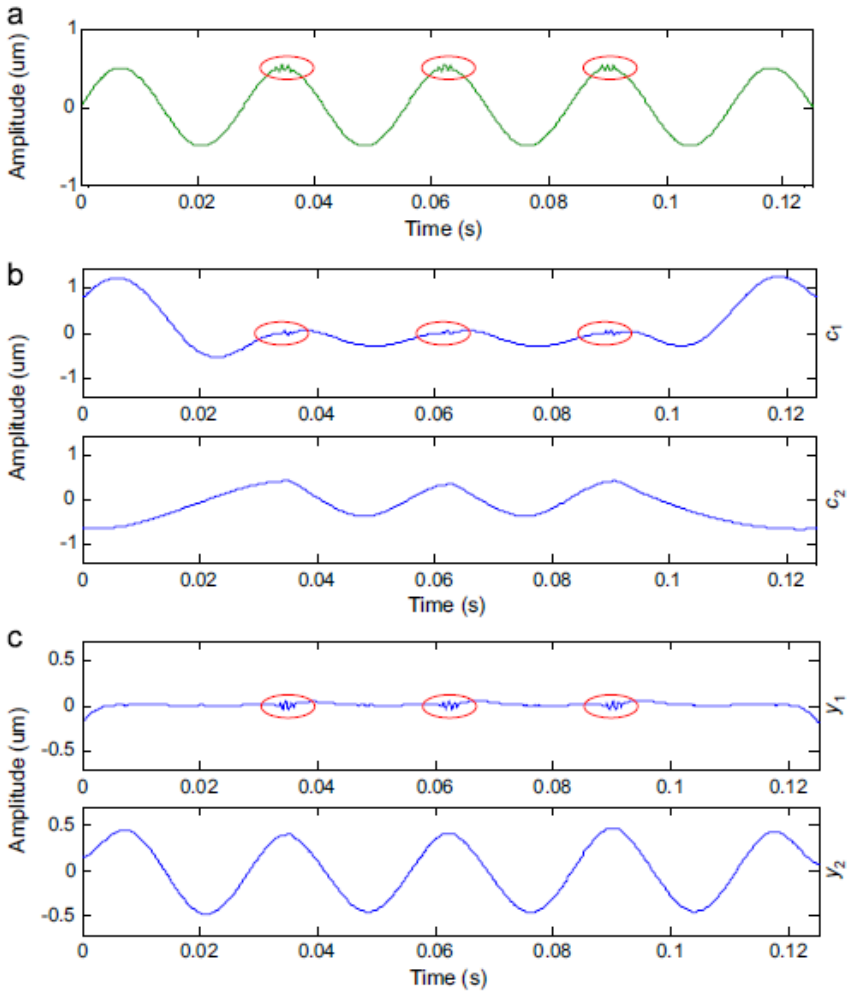


Figure 2.10- A simulated two component signal (a), the decomposed IMFs by EMD (b) and the IMFs decomposed by EEMD (c) (Taken from [167]).

However, another challenge still exists: how to better identify the two effective parameters (the amplitude of added noise and the number of ensemble trials), which affect the performance of the EEMD.

The number of ensemble for EEMD

The effect of the added white noise should decrease following the well-established statistical rule [171]:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}} \tag{2.39}$$

$$\ln \varepsilon_n + \frac{\varepsilon}{2} \ln N = 0 \tag{2.40}$$

where N is the number of ensemble members, ε is the amplitude of the added noise, and ε_n is the final standard deviation of error, which is defined as the difference between the input signal and the corresponding IMF(s).

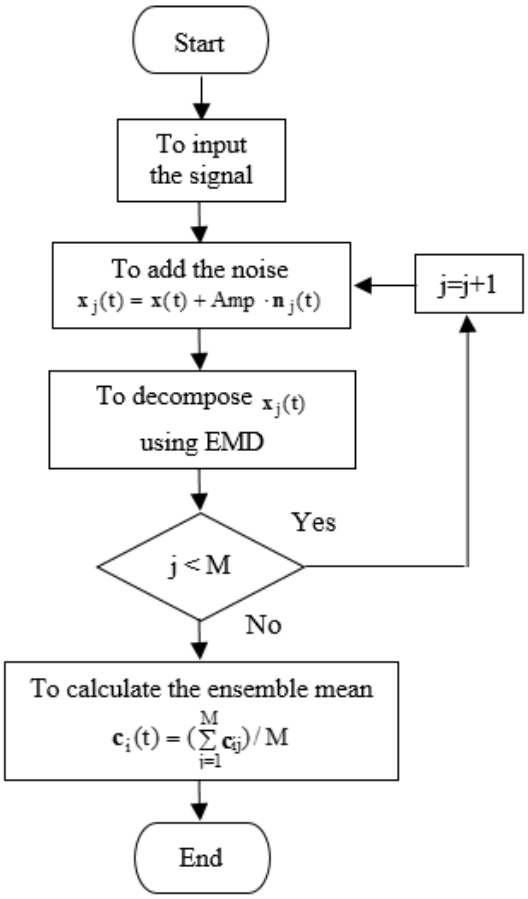


Figure 2.11- The Flowchart of the Ensemble empirical mode decomposition (EEMD) algorithm

In general, an ensemble number of a few hundred will lead to a very good result, and the remaining noise would cause only less than a fraction of 1% of error if the added noise has an amplitude that is a fraction of the standard deviation of the original data [171].

Application

Since the EMD and EEMD are powerful for processing nonlinear and non-stationary signals, these techniques have been studied and widely applied in fault diagnosis of rotating machinery [167]. The following paper review is proposed in terms of different categories called the original EMD method, improved EMD methods, EMD combined with other techniques and EEMD. The studies have been proposed an improvement in EMD, combination of the EMD with other methods to enhance the results and have applied the EEMD for fault detection of roller bearing discussed in this section.

Several improvement have been proposed to enhance the performance of EMD in bearing fault diagnosis. Du and Yang [173,174] improved the local mean calculation of EMD (the conventional envelope mean is substituted by the average mean) and obtained better result in bearing fault diagnosis. Dong et al. [175] improved the efficiency of the sifting process, in which only one time of cubic spline fitting is required in each sifting process. As a result, the technique decreases the computational cost. Terrien et al. [176] presented an algorithm for IMF automatic selection. This algorithm is based on statistical analysis describing the noise repartition between IMFs. Yan and Gao [177] proposed two criteria to determine the most representative IMFs of EMD to recognize the existence and the location of defects in bearings. The improvements in stop criteria of sifting process is discussed in chapter 5.

Many researchers have applied the EMD technique combined with other techniques to achieved better diagnosis results in bearing fault diagnosis. Some of those studies are introduced in this section. Yu et al. [178] and Chen et al. [179] applied EMD and Hilbert-Huang transform to the vibration signals of bearings to create the local Hilbert marginal spectrum for damage detection. Li and Zheng [180] applied the TKEO (Teager-Kaiser energy operator) to decomposed IMFs to calculate instantaneous amplitude. Then, transformed it using the Fourier transform to obtain envelope spectrum, in which the characteristic frequency of the bearing faults can be easily recognized. Rai and Mohanty [181] applied the fast Fourier transform to IMFs to recognize the defective bearing through the characteristic defect frequencies. Li et al. [182] applied Wigner–Ville distribution based on EMD to bearing fault diagnosis and therefore prevented the presence of cross terms. Li et al. [183] developed a method based on EMD, order tracking and Teager-Kaiser energy operator for bearing diagnosis under the run-up or run-down process. Peng et al. [184] applied wavelet packet transform to improve the deficiencies of HHT and formed an improved HHT method for bearing fault detection. Yang et al. [185] used the characteristic amplitude ratios of IMFs as the feature vectors of support vector machines. Cheng et al. [186] build an autoregressive model (AR) for each IMF to recognize the bearing faults. Yang et al. [187] used the energy entropy of EMD as the input of artificial neural networks for identifying bearing damages. Cheng et al. [188] extracted feature vectors applying the EMD and singular value decomposition, then utilized the support vector machines to classify the features of bearings and gears. Lei et al. [189-191] decomposed bearing vibration signals by the EMD to identify fault types and damage sizes using neural networks and genetic algorithms.

The EEMD technique has been used in a number of researches to identify the state of bearings, which is addressed particularly in this section. An et al. [192] applied the Hilbert-Huang transform based on the EEMD to detect the bearing pedestal looseness fault in a wind turbine. Ai et al. [193] presented an approach based on EEMD and envelope spectrum to reliably diagnose bearing defects. Zvokelj et al. [194] combined EEMD and non-linear kernel principal component analysis (PCA) to detect defects of large-size and low-speed bearings. Zhang et al. [195] investigated parameter selection issues (amplitude of added noise and ensemble trial numbers) of EEMD and proposed a modified EEMD method in diagnosing bearing faults. Lu and Wang [196] introduced a method based on EEMD and redundant second generation wavelets denoising to improve the accuracy of bearing fault diagnosis. Lei et al. [197] combined the EEMD and wavelet neural networks (WNN) to construct an automatic fault identification method of locomotive bearings. They used a kurtosis based method to select the sensitive IMFs decomposed by EEMD. Then, the frequency spectrum and envelope spectrum of the selected IMFs were used as input of the WNN. Guo and Tse [198] applied EEMD to bearing fault diagnosis and discussed the influence of parameter setting in EEMD on the results of reducing mode mixing problem.

2.2.3 Teager-Kaiser energy operator

The Teager-Kaiser energy operator (TKEO) was first proposed by Teager [199] and further investigated by Kaiser [200]. A number of studies have been carried out on the theory and applications of the TKEO operator such as [94-97]. In continuous time, the TKEO operator is defined as:

$$\Psi[x(t)] = A^2 = \dot{x}^2(t) - x(t)\,\ddot{x}(t) \tag{2.41}$$

For a simple model, a mass m suspended by a spring with constant k, the displacement measured from its equilibrium position:

$$x(t) = A\ \cos(\omega t + \theta_0)$$

(2.42)

where A is the amplitude of oscillation, ω is the natural frequency ($\omega = \sqrt{k/m}$) and is θ_0 the initial phase.

The total energy of the system is given as the sum of the potential energy of the spring and the kinetic energy of the mass:

$$\mathrm{E} = \frac{1}{2}\ kx^2 + \frac{1}{2}\ m\dot{x}^2$$

(2.43)

It is represented as:

$$\mathrm{E} = \frac{1}{2}\ m\omega^2 A^2$$

(2.44)

It indicates that the total mechanical energy of the simple harmonic oscillation is proportional to the square of the amplitude and the square of the frequency. Applying the TKEO to $x(t)$ gives:

$$\begin{aligned}\Psi[x(t)] = A^2 &= (-A\omega\ \sin(\omega t))^2 - A\ \cos(\omega t)\ (-\omega^2 A\ \cos(\omega t)) \\ &= A^2\omega^2(\sin^2(\omega t) + \cos^2(\omega t))\end{aligned}$$

(2.45)

And finally:

$$\Psi[x(t)] = A^2\omega^2$$

(2.46)

So, the TKEO operator tracks the total energy of the source generating the vibration signal $x(t)$ (per half-unit mass).

For a discrete time signal $x(n)$ (where n is the discrete time index), using difference to approximate differential, the TKEO operator can be formulated as [94]:

$$\Psi[x(n)] = [x(n)]^2 - x(n-1)x(n+1)$$

(2.47)

As at any instant, only three consecutive samples are needed to estimate the instantaneous TKEO, it is adaptive to the instantaneous changes in signals and is quite adapt to resolve transient events. Therefore, it has a good adaptability to the instantaneous changes in signals and an excellent ability to resolve transient events.

It is an adaptive method and effective in estimating the instantaneous frequency and envelope amplitude of nonstationary signals. It has some merits such as low computational cost, high resolution of time and frequency and adaptability to instantaneous feature [98].

The instantaneous frequency and instantaneous amplitude at any time instant of the signal $x(n)$ are defined as:

$$f(n) = \frac{1}{2\pi}\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}}$$

(2.48)

$$|a(n)| = \frac{\Psi[x(t)]}{\sqrt{\Psi[x(t)]}}$$

(2.49)

They can be represented as follows:

$$f(n) = \frac{1}{2}\,arccos\left(1 - \frac{\Psi[x(n+1) - x(n-1)]}{2\Psi[x(n)]}\right)$$

(2.50)

$$|a(n)| = \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1) - x(n-1)]}}$$

(2.51)

In this method, the signal is decomposed into some IMFs by the EMD algorithm. Then, the envelope amplitude and instantaneous frequency of the IMFs are obtained utilizing the TKEO method.

As the TKEO operator is an energy operator, it is expected to be a positive quantity. This issue was first addressed by Maragos et al. [94]. For the continuous version of the TKEO operator, Eq.41, the only way to make the right hand side less than zero is:

$$\dot{x}^2(t) \leq x(t)\ddot{x}(t)$$

(2.52)

Thus, the operator is nonnegative, regardless of the value of $\dot{x}(t)$, if any of the following conditions is satisfied:

- $x(t) = 0$
- $\ddot{x}(t) = 0$
- $x(t) < 0$ and $\ddot{x}(t) > 0$
- $x(t) > 0$ and $\ddot{x}(t) < 0$

This phenomenon was discussed in [201]. As it can be seen in Fig. 2.12 the TKEO of a signal is negative for some values. Based on the definition, the TKEO operator tries to model the energy of the source of the signal, and not the actual measured signal. The signal, in Fig. 2.12, is constructed using two simple sinusoidal signals, where one has considerably higher frequency, but less amplitude than the other signal. Thus, it is made of two separate sources, one high and one low frequency source, where the high frequency source is placed farther away than the low frequency source. In this example, it was incorrectly assumed that only one source is generating the signal that is why a negative value was achieved.

Application

As the TKEO detects a sudden change of the energy stream without a priori assumption of the data structure, it can be utilized for vibration based condition monitoring (non-stationary signals). Junsheng et al. [203] applied the TKEO to each IMFs decomposed by the EMD to extract the instantaneous amplitudes and frequencies. Then, envelope spectra were obtained using the spectrum analysis to look for characteristic frequencies of damaged roller bearings. Li et al. [204] applied the TKEO to the original vibration signals and characteristic frequencies were extracted from envelope spectra. They also implemented a novel method [205] to recognize faults of roller bearing based on Teager-Huang transform (THT). Feng et al. [206] utilized the Fourier spectrum of Teager energy to identify the characteristic frequency of faulty bearings. Liu et al. [207] presented an approach to bearing fault diagnosis based on the TKEO and the Elman neural network. The wavelet packet was used to reduce noise existing in the Teager energy signal, and then feature vectors were extracted from the Teager spectrum. Rodriguez et al. [208] transformed the vibration signal to the Teager-Kaiser domain and featured it with statistical and energy-based measures. The diagnosis was performed with the neural network and the least square support vector machine (LS-SVM). Kwak et al. [209] applied the TEKO in a combination with minimum entropy deconvolution (MED) to detect a defective roller bearing in terms of Kurtosis.

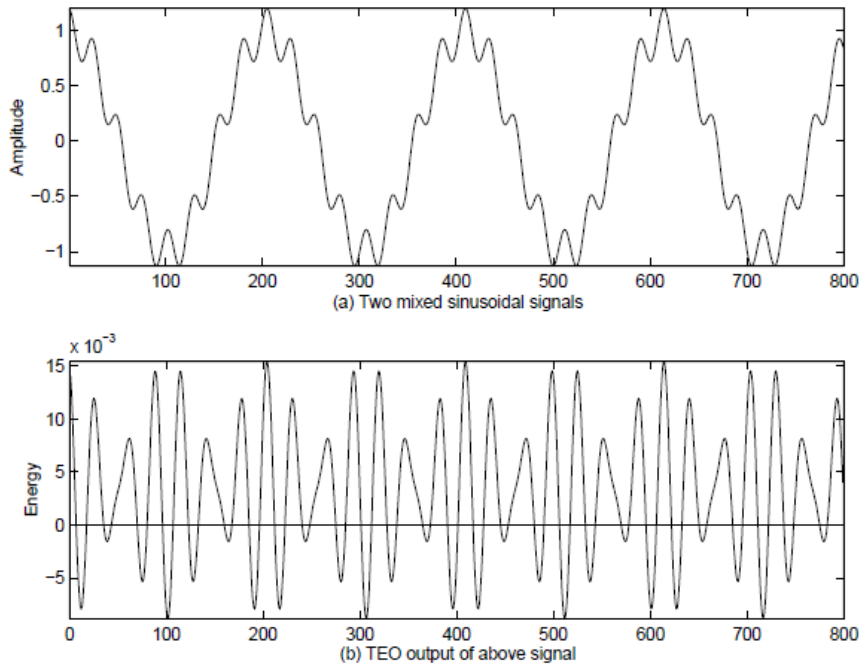


Figure 2.12- the two component signal and the Teager-Kaiser energy operator [202]

2.2.4 Cointegration

The idea of the cointegration is to find within a signal linear combinations of signal segments which are stationary. Engle and Granger [210] formulated one of the first test of cointegration. Johanson proposed a maximum likelihood approach for finding stationary linear combinations of nonstationary variables, which have the same order of integration. [211]. A signal/stochastic process y is said to be integrated of order d ($\boldsymbol{y_t \sim I(d)}$), if it becomes stationary after d times

differencing. Thus, the time series $\boldsymbol{y_t}$ will be integrated of order 1, if its first difference is stationary ($\Delta \boldsymbol{y_t} \sim \boldsymbol{I}(\mathbf{0})$). The order of integration can be estimated by a stationarity test applied on the residuals of the segments, like e.g. the Augmented Dickey-Fuller (ADF) test [212].

Two or more non-stationary time series are said to be cointegrated if a linear combination of them is stationary. The signal Y_t is made of n non-stationary segments/time series ($y_{1t}, \dots, y_{nt}$), which are I (1) integrated.

$$Y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{pmatrix} \tag{2.53}$$

The series is said to be cointegrated with r cointegrating vectors (0 < r < n) if there exists a (n×r) matrix β such that the cointegrating residual u_t is 0 integrated ($u_t \sim I(0)$).

$$\beta' Y_t = \begin{pmatrix} \beta_1' Y_t \\ \vdots \\ \beta_r' Y_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \tag{2.54}$$

The cointegration test based on Johansen procedure [211], tests for the existence of r, $0 \leq r < n$, cointegrating vectors ($\beta_1, \dots, \beta_r$). It determines which of the cointegrating vectors creates the most stationary linear combination.

The vector autoregressive model (VAR) for the signal Y_t is formulated as follows:

$$Y_t = \Psi D_t + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \epsilon_t \tag{2.55}$$

where D_t is deterministic term, p denotes the model order, Ψ and ($\Pi_1, \dots, \Pi_p$) are time-invariant matrices and ϵ_t is error with mean zero and finite variance (i.i.d. $(0, \Omega)$).

The vector error correction model (VECM) takes the form:

$$\Delta Y_t = \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-1} + \Psi D_t + \epsilon_t \tag{2.56}$$

where p denotes the model order, $\Pi = \Pi_1 + \cdots + \Pi_p - I_n$ and $\Gamma_i = -\sum_{j=k+1}^{p} \Pi_j$ (k = 1,…, p - 1).

As Y_t is integrated of order 1, ΔY_t and its lags are stationary, which means that the term ΠY_{t-1} must be I (0). Thus, If Y_t is not cointegrated, the parameter matrix must be zero ($\Pi = 0$) and the VECM model reduces to VAR (p - 1) model in first differences. If the Π matrix have a reduced rank; that is 0 < rank (Π) = r < n, Y_t has r linearly independent cointegrating vectors. The parameter matrix can be decomposed into two n × r matrices:

$$\Pi = \alpha\beta'$$

(2.57)

where α is called the adjustment parameters matrix.

The model (Eq. 2.56) can be formed as:

$$\Delta Y_t = \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \alpha\, \beta' Y_{t-1} + \Psi D_t + \epsilon_t$$

(2.58)

In order to find the cointegrating vectors, first we consider the residuals of the following regressions:

$$\Delta Y_t = \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \widehat{U}_t$$

(2.59)

$$Y_t = \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \widehat{V}_t$$

(2.60)

where $\widehat{U}_t$ and $\widehat{V}_t$ denote the residuals and Φ_1 and Φ_2 are multiplication matrices.

Then, the cointegrating vectors are determined as eigenvectors of the following eigenvalue problem:

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$$

(2.61)

where $S_{00} = \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t \widehat{U}_t'$, $S_{01} = \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t \widehat{V}_t'$, $S_{10} = \frac{1}{T}\sum_{t=1}^{T} \widehat{V}_t \widehat{U}_t'$, $S_{11} = \frac{1}{T}\sum_{t=1}^{T} \widehat{V}_t \widehat{V}_t'$ are the sample covariance matrices.

The eigenvector corresponding to the largest eigenvalue is the most stationary cointegration vector. A likelihood ratio (LR) is utilized to indicate the number of possible cointegrating vectors. The trace statistic (Eq. 2.62) is utilized to test the hypothesis that there are at most r cointegrating vectors.

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$$

(2.62)

where $\hat{\lambda}_i$ denotes the estimated eigenvalues of the matrix Π. If the trace test is higher than the critical value, tabulated in [211], the null hypothesis is rejected.

First the null hypothesis ($H_0(r_0 = 0)$) is tested against the alternative hypothesis ($H_1(r_0 > 0)$). If the null is not rejected then there does not exist cointegrating vectors. If the null is

rejected, there is at least one cointegrating vector and the test is continued to test $H_0(r_0 = 1)$ against $H_1(r_0 > 1)$. There exists only one cointegrating vector, if the null is not rejected. Otherwise, it is concluded that there is at least two cointegrating vectors. The procedure is continued until the null is not rejected.

Application

Applications of cointegration to finance may be found in [213-217]. Cointegration has been applied recently in structural health monitoring (SHM) as a powerful method to remove nonstationarity produced by environmental variations such as temperature, wind and humidity. Cross and Worden [218] discussed why cointegration is applicable to engineering data .Cross et al. [219] tested successfully the idea of using cointegration for SHM. It could detect the damage introduced in a composite plate. Antoniadou et al. [220] applied Hilbert-Huang and Teager- Kaiser transforms to extract relevant information and used cointegration to remove non-stationarity produced by variation of environment conditions. Worden et al. [221] demonstrated how a multiresolution approach to cointegration can enhance sensitivity for damage detection.

Chapter 3

Pattern recognition

In this chapter pattern processing (recognition) stage is described. Pattern recognition (PR) is the stage where a method is applied to interpret the damage state based on the given feature vectors extracted by signal processing techniques. PR system is trained using labelled training data in all those methods called supervised learning. However, when the available data are not labelled, unsupervised algorithms are used to recognize previously unknown patterns. As described in section 1.3.5, depending on the desired diagnosis, the PR algorithms can be categorized in two groups; Novelty detection, which uses unsupervised learning, and classification, which uses supervised learning algorithm. Nearest neighbour classifier, Artificial neural networks (ANNs) and Support vector machine (SVM) are the classification techniques usually used for fault diagnosis of rotating machines described in details in this chapter. As a novelty detection method, one-class SVM is introduced in section 3.3.3.

3.1 Nearest-neighbour classifier

Nearest neighbour classifier is a non-parametric pattern recognition method, which means that it does not make any assumption on the statistical properties of data. The simple idea is to find the closest neighbour for new given data among the training data and assign the label of training data to new data. It is usually performed by evaluating a dissimilarity function or distance between the training data and new samples [222]. The Euclidean distance usually used as a measure of neighbourhood can be formed as follows:

$$d(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{k=1}^{n} (x_k - x'_k)^2} \tag{3.1}$$

where x denotes the training data including n attributes $(x_1, \dots, x_n)$ and x' is the new given data $(x'_1, \dots, x'_n)$.

The distance introduced can be generalized using the Minkowski distance metric with the parameter r:

$$d_r(\boldsymbol{x}, \boldsymbol{x}') = \left(\sum_{k=1}^{n} (x_k - x'_k)^r \right)^{1/r} \tag{3.2}$$

The global distance can be defined as a weighted sum of the local distances. It allows to consider different importance for different attributes in the computation of the overall distance.

$$d_w(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{k=1}^{n} w_k \, (x_k - x'_k)^2} \tag{3.3}$$

where w_k denotes the weight for the k-th dimension. The weights can be set between zero and one; a weight of zero indicates a totally irrelevant attribute.

A weighted average distance is formulated as follows:

$$d_w(\boldsymbol{x}, \boldsymbol{x}') = \frac{(\sum_{k=1}^{n} w_k \, (x_k - x_k')^2)^{1/2}}{\sum_{k=1}^{n} w_k} \tag{3.4}$$

The distance introduced has the following properties:

- Non-negativity: $d(\boldsymbol{x}, \boldsymbol{x}') \geq 0$ and $d(\boldsymbol{x}, \boldsymbol{x}) = 0$
- Symmetry: $d(\boldsymbol{x}, \boldsymbol{x}') = d(\boldsymbol{x}', \boldsymbol{x})$
- Triangular inequality: $d(\boldsymbol{x}, \boldsymbol{x}') \leq d(\boldsymbol{x}, \boldsymbol{x}'') + d(\boldsymbol{x}'', \boldsymbol{x}')$

When a measure satisfies all three properties it is known as metric.

K-Nearest Neighbour method (KNN)

The K-Nearest Neighbour classifier finds K nearest training points of $\boldsymbol{x}'$ and uses the majority vote of their labels to assign a label to $\boldsymbol{x}'$. There exist two common voting approaches; majority voting and inverse distance-weighted voting. For majority voting approach, all votes are equal we count how many of the K neighbours have that class. Then, the class with the most votes is returned. This approach is shown in Fig. 3.1 for the 1-, 2- and 3-nearest neighbours of a data point located at the centre of each circle. Following the procedure described, the data point in Fig. 3.1a is classified as negative, while in Fig. 3.1c is labelled as positive. The situation shown in Fig. 3.1b is the state that there is a tie between the classes and the final choice is taken through a random choice. A drawback of this approach occurs when the class distribution is skewed. That is, examples of a more frequent class tend to dominate the prediction of the new example, because they tend to be common among the K nearest

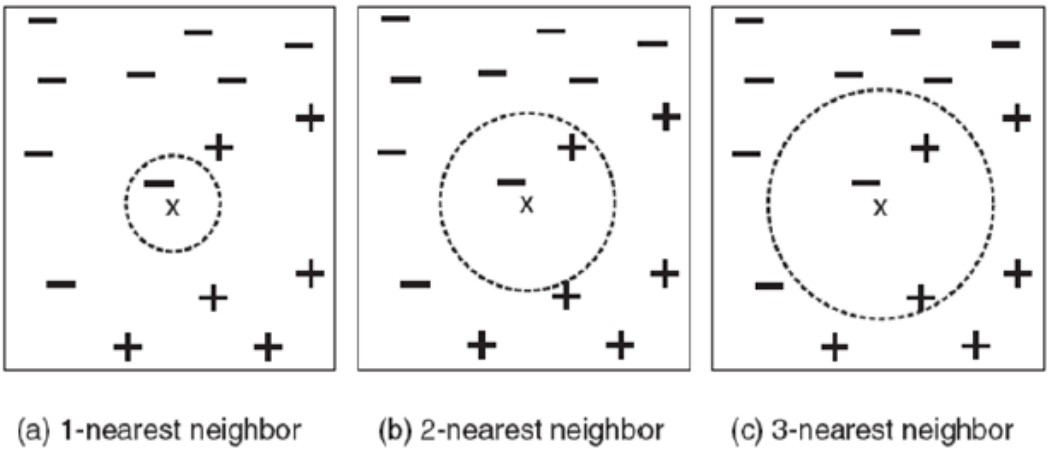


Figure 3.1- The 1-, 2- and 3-nearest neighbours of an instance (from [222]).

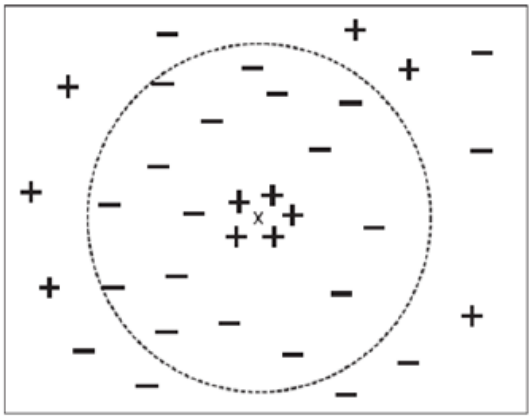


Figure 3.2- The K-nearest neighbours with large value of K (from [222]).

neighbours due to their large number [223]. One solution is to weight the classification so that closer neighbors get higher votes. The class of each of the K nearest points is multiplied by a weight proportional to the inverse of the distance from that point to the test point. Then, the votes are summed and return the class with the highest vote.

Although larger values of K (shown in Fig. 3.2) reduce the effect of noise on the classification [224], it makes boundaries between classes less distinct. For two class classification problems, K is chose to be an odd number to avoid tied votes.

Application

The Nearest neighbour classifier has been used in a number of fault diagnosis studies as a simple PR method [23]. Mechefske and Mathew [225] used the Nearest-Neighbour classifier for automatic fault diagnosis in low speed ($\leqslant$100 RPM) rolling element bearings. Zhang et al. [226] used genetic programming is used for feature generation, and then the K-Nearest Neighbour is applied as a classifier on roller bearing data. Trendafilova [227] applied a combination of the modified Principal component analysis (PCA) and the One-Nearest Neighbour approach to detect fault of a ball bearing automatically and classify to one of the considered fault categories. He et al. [228] applied HHT to fault identification of full ceramic bearing and compared the performance of the KNN based fault classifier with a neural network based classifier. Jiang et al. [229] applied KNN classifier to recognize different fault categories and severities of bearings. Jiang et al. [230] proposed a new method based on Marginal Fisher analysis to excavate low-dimensional fault characteristics embedded in the high-dimensional feature space and fed into the KNN classifier to recognize different fault categories.

3.2 Artificial neural networks (ANNs)

The basic processing unit of the brain is the neuron, which acts by summing inputs from connected neurons. If the total summation or activation exceeds a certain threshold, the neuron produce a single output or response [29]. The Artificial neural network (ANNs) attempts to find mathematical representations of information processing in biological systems [231]. The basic neural network model is constructed by functional transformations. First, linear combinations of the input variables (x_i) are formed as follows [231]:

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \qquad j = 1, \dots, M \tag{3.5}$$

where parameters $w_{ji}^{(1)}$ are weights, $w_{j0}^{(1)}$ denote biases and quantities a_j are known as activations. The superscript (1) denotes that the parameters correspond to the first layer of the network.

Then, the activations are transformed to find *hidden unit* shown in Fig. 3.3:

$$z_j = h(a_j) \tag{3.6}$$

The hyperbolic tangent and sigmoid functions are usually used as activation functions [29, 231]. The linear combination of the hidden units are used to obtain output unit activations, which corresponds to the second layer of the network:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \qquad k = 1, \dots, K \tag{3.7}$$

where K is the total number of outputs, $w_{k0}^{(2)}$ and $w_{kj}^{(2)}$ are bias parameters and weight parameters corresponding to the second layer, respectively.

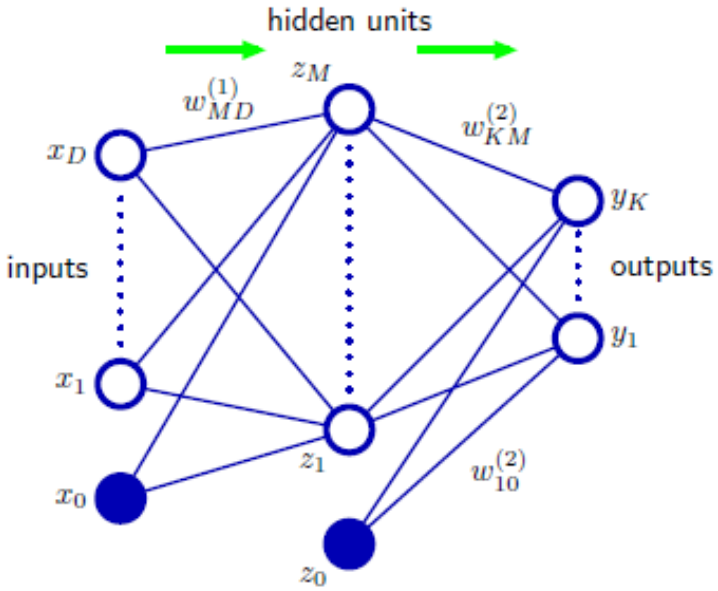


Figure 3.3- Two-layer neural network (single-hidden-layer network) (from [231]).

Finally, the output unit activations are transformed to obtain network outputs y_k.

$$y_k = \sigma(a_k) \tag{3.8}$$

The overall network function described in Eq. 3.9, shows that the neural network model is a nonlinear function from a set of input variables x_i to a set of output variables y_k controlled by adjustable parameters w (all weight and bias parameters) [231].

$$y_k(x, w) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)}\, h\left(\sum_{i=1}^{D} w_{ji}^{(1)}\, x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right) \tag{3.9}$$

Multi-layer network is a network built up with a number of hidden layers. The output of the node i in the layer m ($x_i^{(m)}$) is obtained as follows:

$$x_i^{(m)} = h\left(z_i^{(m)}\right) = h\left(\sum_j w_{ji}^{(m)}\; x_i^{(m-1)}\right) \tag{3.10}$$

where $x_i^{(m-1)}$ denotes all signals from the preceding layer $m-1$.

Before using a network, the weight parameters must be adjusted in the training phase. Supervised learning is the training algorithm usually used. A set of known network inputs and outputs are used so that at each training step, a set of inputs is passed forward through the network to obtain estimated outputs $\hat{y}_i$ and compare with the desired outputs y_i. If the comparison error is considered small enough, the weights does not require any adjustment. Otherwise, the error is used to adjust the weight parameters so that the error is reduced. For each presentation of a training set, the network error is evaluated as follows [29]:

$$J(t) = \frac{1}{2}\sum_{i=1}^{n^{(l)}}\left(y_i(t) - \hat{y}_i(t)\right)^2 \tag{3.11}$$

where J is a function of the network parameters $J(w)$, $n^{(l)}$ is the number of output layer nodes and the integer t labels the presentation order of the training sets.

Generalization problem or overfitting occurs when there exist too many parameters in the model compared to the number of training points. In this case, when the model is applied to different data collected from the same system, the prediction results could be unsatisfactory and the model fails to generalize. In order to avoid overfitting, cross-validation on an independent dataset can be used. The data is divided into training, validation and testing sets. The network is trained on the training data and the error is calculated on the validation data. When the error begins to rise even though the error on the training set continues to decrease, the number of hidden units has reached the point where overfitting is beginning. The number

of hidden units can be fixed at the point where the minimum error on the validation set occurred [29]. Then, generalization can be examined using independent testing set.

Application

A number of applications of neural network have been performed for condition monitoring of structures, rotating machines and bearings in particular. A review of a variety of diagnostic techniques based on artificial neural approach for rotating machinery was presented by Yang et al. [232]. McCormick and Nandi [233] used multi-layer network to classify the condition of a small rotating machine. The time/frequency-domain analysis and neural networks were applied for motor rolling bearing fault diagnosis by Li et al. [234]. Vibration simulations and experimental results obtained indicate that neural networks can effectively identify various motor bearing faults. Samanta et al. [235] compared the performance of three types of artificial neural networks, namely, multilayer perceptron (MLP), radial basis function (RBF) network, and probabilistic neural network (PNN) for bearing fault detection. The results showed that the classifiers are effective in detection of the bearing condition. Tyagi [236] used simple statistical features such as standard deviation, skewness, kurtosis etc. of the time-domain vibration signal of rolling element bearing along with peaks of the signal and peak of power spectral density (PSD) as features to compare the performance of the ANN and SVM classifier. Jayaswal et al. [237] provided a review of recent developments in the area of applications of ANN and Wavelet Transform in fault diagnosis. Zarei [238] showed that using time domain features and ANN can be effective in accurate diagnosis of various motor bearing faults with high precision.

3.3 Support vector machine (SVM)

Classical learning methods (like Neural Networks) attempts to minimize error on the training data set called the empirical risk minimization (ERM). On the other side, Support vector machine (SVM) is designed based on the structural risk minimization (SRM) principle. It provides a trade-off between the complexity of the model used for fitting and the quality of fitting the training data. It gives better generalization abilities and SRM is achieved through a minimization of the upper bound of the generalization error [239-241].

The three types of SVM (Two-class, Multi-class and One-class) and their fundamentals are outlined as follows.

3.3.1 Two-class SVM

Linear SVM is the simplest case that the data which belongs to two different classes can be separated using a linear hyperplane, as it can be seen in Fig. 3.4. Consider the data $\boldsymbol{x}_i$ (i=1, ..., n) labelled as positive ($y_i = +1$) or negative classes ($y_i = -1$). The separating line is defined as

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{3.12}$$

where the parameters $\boldsymbol{w}$ and b are the normal vector and scalar, respectively.

All the training data satisfy the following constraints:

$$\boldsymbol{w} \cdot \boldsymbol{x} + b \geq +1 \text{ if } y_i = +1$$

$$\boldsymbol{w} \cdot \boldsymbol{x} + b \leq -1 \text{ if } y_i = -1 \tag{3.13}$$

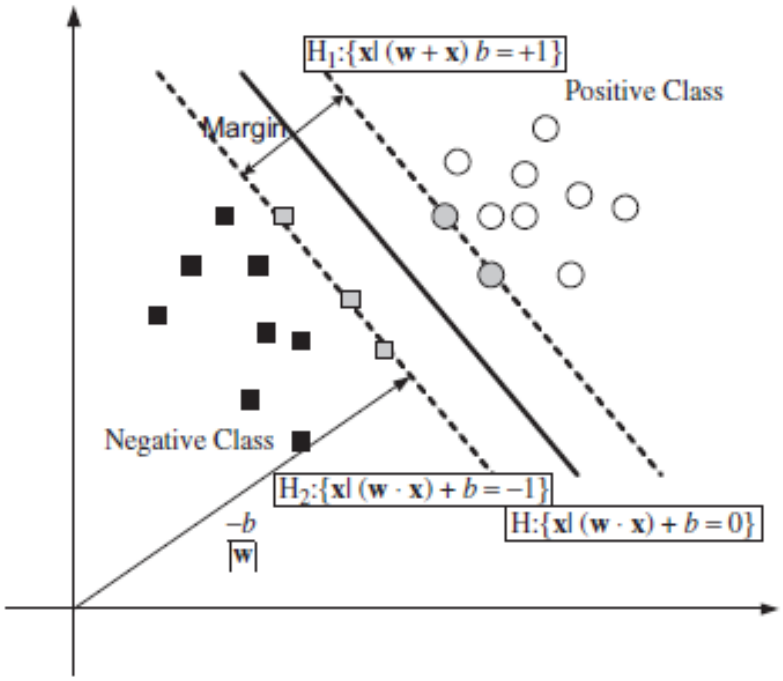


Figure 3.4- Classification of two classes of data using SVM (from [116])

It can be represented as follows:

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x} + b) \geq 1 \quad \forall i \tag{3.14}$$

As it can be seen in Fig. 3.4, the margin is defined as:

$$margin = 2/\|\boldsymbol{w}\| \tag{3.15}$$

The constructed hyperplane, which creates the maximum margin, is called the optimal separating hyperplane. It can be formulized as the following optimization problem:

$$\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2\right) \tag{3.16}$$

In real problems, an exact line dividing the data is usually difficult to obtain so that, by introducing the slack variable (ξ_i) (shown in Fig. 3.5) and ignoring few outlier data points, a smooth boundary can be created as well as the model called soft margin-SVM.

The Eq. 3.16 is modified as follows:

$$\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right)$$

$$\text{subject to } \begin{cases} y_i(\boldsymbol{w}\cdot\boldsymbol{x}+b) \geq 1-\xi_i \\ \xi_i \geq 0 \end{cases} \quad i = 1,\ldots,N \tag{3.17}$$

where ξ_i is measuring the distance between the hyperplane and the samples laying in the wrong side of the hyperplane whilst C is the error penalty.

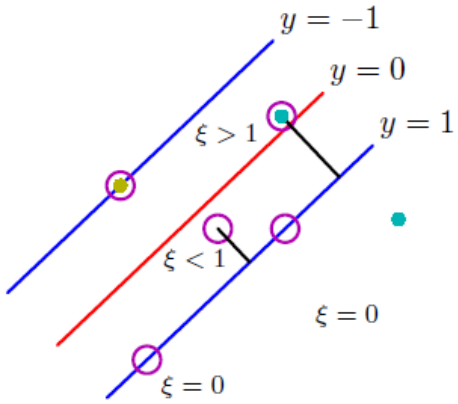


Figure 3.5- The slack variables ξ_i and support vectors (data points with circles around them) (from [231]).

Using the Kuhn-Tucker condition the problem is converted into the equivalent Lagrangian dual problem [231]

$$\min L(\boldsymbol{w},b,\alpha) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{N}\alpha_i y_i\,(\boldsymbol{w}\cdot\boldsymbol{x}+b) + \sum_{i=1}^{N}\alpha_i \tag{3.18}$$

where α_i are the Lagrangian multipliers. Now, in order to minimize L with respect to $\boldsymbol{w}$, b the saddle point equations are formed as:

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0, \qquad \frac{\partial L}{\partial b} = 0$$

The result implies that the solution can be expressed as a linear combination of the training vectors:

$$\boldsymbol{w} = \sum_{i=1}^{N}\alpha_i y_i \boldsymbol{x}_i\,, \qquad \sum_{i=1}^{N}\alpha_i y_i = 0 \tag{3.19}$$

Moreover, b can be obtained from any support vector $\boldsymbol{x}_i$, which lies on the margin and satisfies Eq. 3.20. Thus the scalar parameter b can be obtained using the support vectors from Eq. 3.21.

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) = 1 \tag{3.20}$$

$$b = y_i - (\boldsymbol{w} \cdot \boldsymbol{x}_i) \tag{3.21}$$

Substituting $\boldsymbol{w}$ and b from Eqs. 3.19 and 3.21, into the optimization problem Eq. 3.18, the dual quadratic optimization problem is obtained:

$$\max L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j \, \boldsymbol{x}_i \cdot \boldsymbol{x}_j$$
$$\text{subject to } \begin{cases} \sum_{i=1}^{N} \alpha_i y_i \\ 0 \le \alpha_i \le C \end{cases} \quad i = 1, \ldots, N \tag{3.22}$$

Solving the optimization problem Eq. 3.22, the Lagrangian multipliers (α_i) are obtained. These multipliers are required to achieve the $\boldsymbol{w}$ from the Eq.19. The examples $\boldsymbol{x}_i$ for which $\alpha_i > 0$ are those points that are on the margin, or within the margin when a soft-margin SVM is used. These are the so-called support vectors.

Non-linear decision function to label new data is:

$$f(\boldsymbol{x}) = sign\left(\sum_{i,j=1}^{N} \alpha_i y_i \left(\boldsymbol{x}_i \cdot \boldsymbol{x}_j \right) + b \right) \tag{3.23}$$

The SVM could also be applied in a case of non-linear classification by mapping the data onto a high dimensional feature space, where the linear classification is hence possible as shown in Fig. 3.6.

A non-linear vector function such as $\boldsymbol{\Phi}(\boldsymbol{x}) = \left(\varphi_1(\boldsymbol{x}), \ldots, \varphi_l(\boldsymbol{x})\right)$ is used to map the n-dimensional input vector x onto l dimensional feature space, so that the decision function becomes

$$f(\boldsymbol{x}) = sign\left(\sum_{i,j=1}^{N} \alpha_i y_i \left(\boldsymbol{\Phi}(\boldsymbol{x}_i) \cdot \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right) \right) + b \right) \tag{3.24}$$

By applying the Kernel function as the inner product of mapping functions Eq. 3.25, it is not necessary to explicitly evaluate mapping in the feature space.

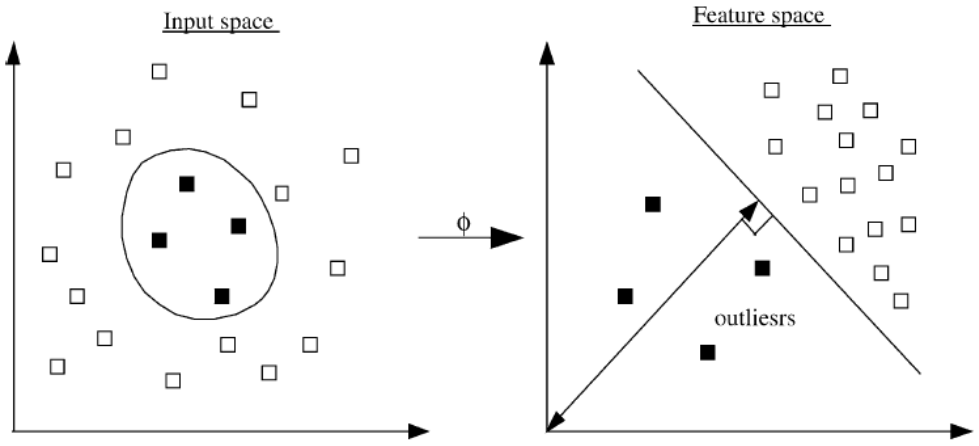


Figure 3.6- Non-linear mapping from input space to high-dimensional feature space (from [242])

$$\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\boldsymbol{\Phi}(\boldsymbol{x}_i) \cdot \boldsymbol{\Phi}(\boldsymbol{x}_j)\right) \tag{3.25}$$

Various kernel functions could be used such as:

- Linear $\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j)^d$
- Polynomial $\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^d$
- Guassian radial basis function (RBF) $\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left(-\boldsymbol{\gamma}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right)$
- Hyperbolic tangent $\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = tanh(k\boldsymbol{x}_i \cdot \boldsymbol{x}_j + c)$

As the kernel function defines the feature space in which the training set is classified, the selection of the appropriate kernel function is very important.
The bias parameter (Eq. 3.21) using support vectors ($\boldsymbol{x}_i$ with $0 < \alpha_i <$ C) will be:

$$b = y_i - \left(\sum_{j=1}^{N} \alpha_j y_j \, \boldsymbol{K}(\boldsymbol{x}_j, \boldsymbol{x}_i)\right) \tag{3.26}$$

Utilizing a kernel function, Non-linear decision function will be:

$$f(\boldsymbol{x}) = sign\left(\sum_{i,j=1}^{N} \alpha_i y_i \, \boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) + b\right) \tag{3.27}$$

The error penalty (C) in Eqs. 3.17 and 3.22, is a parameter that allows one to trade off training error versus model complexity. A small value for C will increase the number of training errors, while a large C will lead to a behavior similar to that of a hard-margin SVM. The best values for the C parameter and kernel parameter are determined by optimising the performance of the classifiers on the validation sets or applying cross-validation for verifying performance using only a training set.

3.3.2 Multi-class SVM

Sometimes, the problems of interest involves more than two classes such as for the case of fault diagnosis of rotating machineries where several fault sources exist. There exist two strategy to construct a multi-class SVM: One-against-all and One-against-one [116].

- One-against-all (OAA)

In this method k different classifier SVM are constructed where k is the number of classes. The training data in the i-th class with positive labels, and all the other data (k-1 classes) with negative labels are utilized for training the i-th SVM. For the given data $\{x_i, y_j\}$ where $x_i \epsilon R^n$, $i = 1, \dots, l$ and $y_j \epsilon \{1, \dots, k\}$ is the label of each class of x_i, the problem of constructing i-th SVM is [116]:

$$min\ \frac{1}{2}\|\boldsymbol{w}^i\|^2 + C\sum_{i=1}^{l}\xi_j^i\left(\boldsymbol{w}^i\right)^T \tag{3.28}$$

$$\text{Subject to } \begin{cases} \left(\boldsymbol{w}^i\right)^T\phi\left(x_j\right) + b^i \geq 1 - \xi_j^i & if \quad y = i \\ \left(\boldsymbol{w}^i\right)^T\phi\left(x_j\right) + b^i \leq -1 + \xi_j^i & if \quad y \neq i \\ \xi_j^i \geq 0 & j = 1, \dots, l \end{cases}$$

where the training data is mapped to a higher-dimensional space by function ϕ.

- One-against-one (OAO)

In this method $k(k-1)/2$ SVM models are constructed so that each model is trained using data from two classes. For training data from the i-th and the j-th classes, the classification problem is formed as follows [116]:

$$min\ \frac{1}{2}\|\boldsymbol{w}^{ij}\|^2 + C\sum_{t}\xi_t^{ij}\left(\boldsymbol{w}^{ij}\right)^T \tag{3.29}$$

$$\text{Subject to } \begin{cases} \left(\boldsymbol{w}^{ij}\right)^T\phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij} & if \quad y = i \\ \left(\boldsymbol{w}^{ij}\right)^T\phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij} & if \quad y \neq i \\ \xi_t^{ij} \geq 0 & j = 1, \dots, l \end{cases}$$

There exist various approaches for future testing after constructing all classifiers. After some tests, the decision is made as follows: if $sign\left(\boldsymbol{w}^{ij}\right)^{T}\phi(x_t)+b^{ij}$ says x is in the i-th class, then the vote for the i-th class is added by one. Otherwise, the j-th is increased by one. Then x is predicted in the class using the largest vote [116].

3.3.3 One-class SVM

One-class SVM constructs the separating hyperplane using only one class of data and to classify new samples belong to other possible classes as outliers [113]. As it can be seen in Fig. 3.7, it constructs a hyperplane around the data, such that its distance to the origin is maximal among all possible hyperplanes. The one-class SVM (like supervised two-class SVM) could also be applied in a case of non-linear classification by mapping the data onto a high dimensional feature space, where the linear classification is then possible.

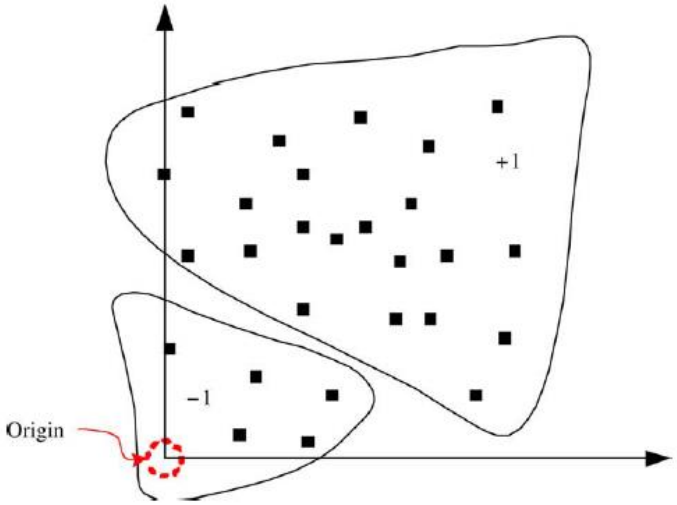


Figure 3.7- Classification by one-class SVM (from [242]).

The Margin is defined as [113]:

$$margin = \rho/\|\boldsymbol{w}\| \tag{3.30}$$

To separate the data set from the origin, the following quadratic program must be solved [113]:

$$\min\left(\frac{1}{2}\|\boldsymbol{w}\|^2+\frac{1}{vl}\sum_{i=1}^{l}\xi_i-\rho\right) \tag{3.31}$$

$$\text{subject to } \begin{cases} y_i\left(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i)\right)\geq\rho-\xi_i \\ \xi_i\geq 0 \end{cases} \quad i=1,\ldots,l$$

where $\boldsymbol{w}$ and ρ are the weight vector and the offset parameterizing the hyperplane. ξ_i is the slack variable, v is the regularization parameter and represents an upper bound on the fraction of outliers (training errors) and a lower bound on the fraction of support vectors (SVs) with respect to the number of training samples. It is a variable taking values between 0 and 1 that monitors the effect of outliers (hardness and softness of the boundary around data).

Introducing Lagrange multipliers we obtain the dual problem as:

$$min\ \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j\,\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)$$

$$\text{Subject to } \begin{cases}0 \le \alpha_i \le \frac{1}{\upsilon l}\\ \sum_{i=1}^{N}\alpha_i = 1\end{cases} \tag{3.32}$$

If υ approaches 0, the upper boundaries on the Lagrange multipliers tend to infinity, so the second inequality constraint in Eq. 3.32 becomes void. As the penalization of errors becomes infinite, it returns to the corresponding hard margin algorithm.

For the positive, non-zero multipliers (support vectors x_i)) we will have:

$$\rho = \boldsymbol{w}\cdot\phi(\boldsymbol{x}_i) = \sum_{j=1}^{l}\alpha_j\,\boldsymbol{K}(\boldsymbol{x}_j,\boldsymbol{x}_i) \tag{3.33}$$

Accordingly the non-linear decision function for labelling new samples is represented as follows:

$$f(\boldsymbol{x}) = sign\left(\sum_{i=1}^{l}\alpha_i\,\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}) - \rho\right) \tag{3.34}$$

Applications

Support vector machine has been widely used for damage detection and identification of rotating machines. Widodo et al. [116] presented a very complete review of the recent developments of SVM in machine condition monitoring and fault diagnosis.

Jack and Nandi [243] calculated statistical features for vibration signals of roller bearings and selected the optimal features using genetic algorithm (GA). They used SVM and ANN for fault identification. Samanta et al. [244,245] proposed an improved method in fault detection of bearing. They applied GA for feature selection and searching proper RBF kernel parameters. Several effect conditions such as sensor location, signal preprocessing, number of features were presented to show the performance of SVM compared with ANN. Shin et al. have presented in [242] a comparison of the results of damage detection for one-class SVM and multi-layer networks for electro-mechanical machinery. Experiments performed on artificial and real dataset showed that the performance of one-class method is mostly superior to that of multilayer perception. Yang et al. [246] proposed a fault diagnosis method based on multi-class SVMs for a rotating machinery. A comparison between the SVMs and ANNs algorithms were performed and the great characteristics of SVMs were demonstrated. Zhang et al. [247] proposed probabilistic SVM to reduce the number of samples for fault diagnosis of bearing. Hu et al. [248] used improved wavelet package transform (IWPT) and SVM ensemble for fault diagnosis of rolling element bearing. Yang et al. [249] defined the ratios of amplitudes at the different fault characteristic frequencies in the envelope spectra of some IMFs (obtained using EMD) as the characteristic amplitude ratios. Then, they were used as input to the support vector machine (SVM) classifiers to identify fault patterns of the roller bearings. Pirra et al. [250]

used feature vectors extracted by EMD as input for one-class SVM to detect defect of roller bearing operating under different speed and external load conditions. Zhang and Zhou [251] applied ensemble empirical mode decomposition (EEMD) and optimized support vector machine (SVM) for multi-fault diagnosis of rolling element bearings.

Chapter 4

Experimental datasets

The fault detection approaches proposed in this thesis have been tested on experimental datasets. In order to validate the methods for early damage detection, the acceleration signals were collected for healthy and faulty bearing using the test rig described in section 4.1. Validation of the method proposed for damage detection of bearing operating under time-varying condition are tested using the data collected on the test rig 2 described in section 4.2. In this chapter the procedure of data acquisition process is presented and the test rigs are described in details in sections 4.1 and 4.2, respectively.

4.1 Test rig1

The test rig 1 shown in Fig. 4.1 has been developed and assembled by the Dynamics & Identification Research Group (DIRG) at the Department of Mechanical and Aerospace Engineering of Politecnico di Torino. It is equipped with an electro-spindle, its power supply, three bearings and their supports (1 and 3 in Fig. 4.1), a load applying mechanism, a load cell, four accelerometers to measure acceleration signals and thermocouples to measure temperature of the oil and bearings. It is possible to change the rotational speed, the external load applied and the oil temperature. The electro-spindle has been selected so that a rotating speed up to 30000 rpm can be reached. The radial force is applied by the green springs to the bearing shown by 3 in Fig. 4.1. Moreover, it is possible to heat the oil in the tank with an electric resistance in order to change its temperature.

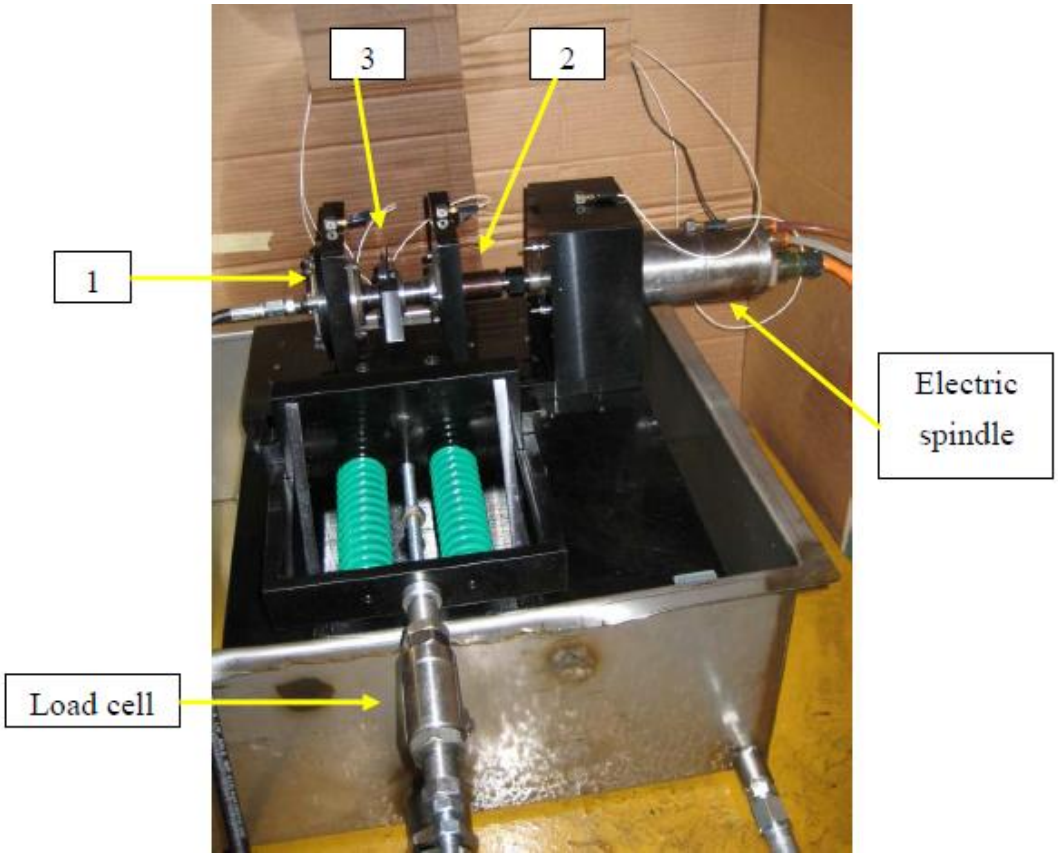


Figure 4.1- The test rig 1 assembled in DIRG group.

4.1.1 Description

The test rig has been equipped with triaxial accelerometers fixed to the bearing supports using a set screw. Their axes orientation x, y and z axis can be seen in Fig. 4.2 and correspond to the axial, radial and tangential direction, respectively. Technical characteristics of the accelerometers are introduced in Table 4.1. Kistler piezoelectric accelerometers utilize ceramic shear sensing elements that generate an electrical charge when mechanically loaded.

The calibration characteristic of the load cell is illustrated in Fig. 4.3, which shows the sensitivity of load cell for measuring the radial load: it is 0.499 mV/N. In order to record some significant temperature values, two thermocouples are placed in the oil basin and in proximity of the external ring of bearing 3, respectively.

All the accelerometers are connected to the OROS OR38 data acquisition system shown in Fig. 4.4. The data acquisition system is able to sample up to 24 channels synchronously with a maximum sampling rate of 102.400 kHz with a passing band of 40 kHz and anti-aliasing filters. It can be set to measure time record histories up to many minutes each channel.

The bearings used in the test rig are from an aeronautical application (type of bearing is SKF NN 76141). SKF has also provided the creation of various damages in their laboratories. A list of artificial defect type and their extent is presented in Table 4.2 and an example of these damages is shown in Fig. 4.5.

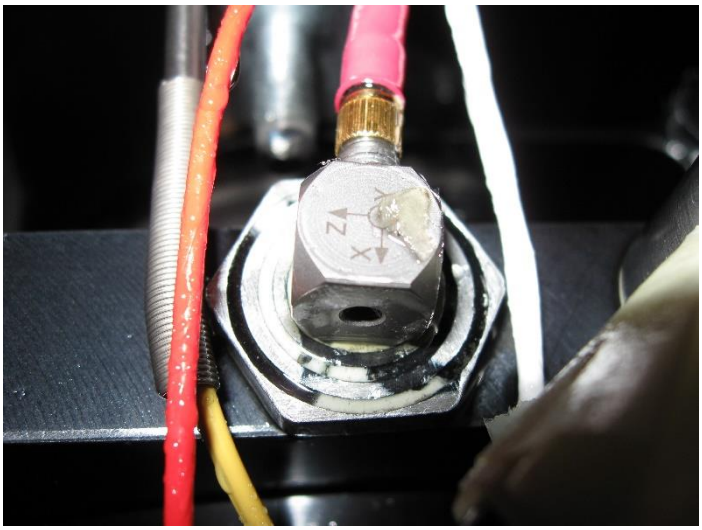


Figure 4.2- The triaxial accelerometer used to collect data in three directions.

Table 4.1- Technical characteristics of the accelerometers

Producer	Model	Full scale	Nominal sensitivity	Resonance
Kistler	8763A500, triaxial	500 g	10 mV/g	~55 kHz

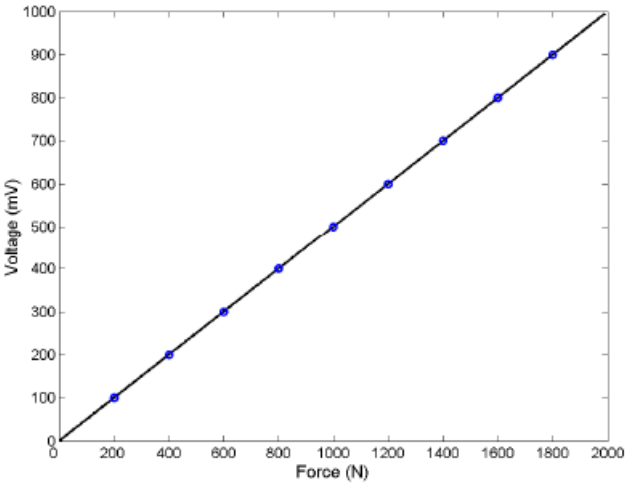


Figure 4.3- The Characteristic of the load cell.

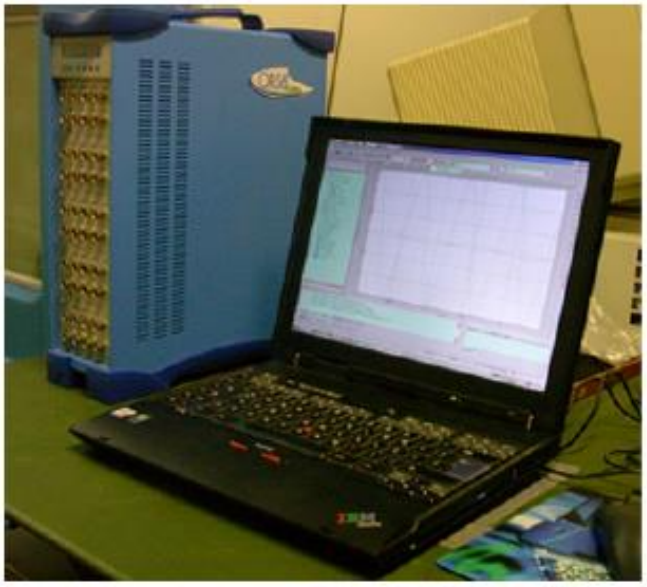
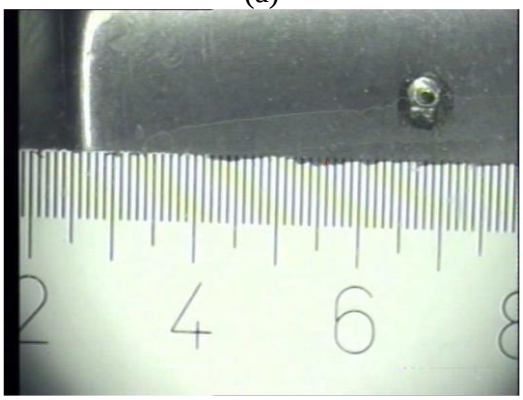


Figure 4.4- Oros OR38 acquisition system with its PC.

Table 4.2- Various type of defects and sizes created on the different bearings

Defect location	Inner ring	Inner ring	Inner ring	Roller element	Roller element	Roller element
Size (microns)	150	250	450	150	250	450

(a)



(b)

Figure 4.5- The defective bearing with damage created on one roller (a) the defect size with the scale in tenths of millimeters (the values 2, 4, 6 and 8 are in millimeters).

The test rig consists of four accelerometers used for the acceleration acquisitions, as shown in Fig. 4.6. Three accelerometers located on the bearing supports in order to collect the vibration signals. They are connected to channels 1-2-3, 4-5-6 and 7-8-9. The last accelerometer, giving channels 10-11-12, is placed on the elector-spindle support in order to acquire the mechanical noisy components given by the spindle itself. The defective bearings are assembled in the support where the accelerometers connected to channels 1-2-3 are mounted. Moreover, two thermocouples used to measure the bearing and oil temperatures. One of them is placed in contact with the support of the bearing and the other is put in the oil tank.

Datasets required to validate the proposed techniques for early damage detection were acquired using the test rig 1. The healthy bearing and six defective bearings described in Table 4.2 were mounted and, for each of them, the tests were performed under five different shaft speeds (100, 200, 300, 400 and 500 Hz) and three values of loading (1.0, 1.4 and 1.8 kN). A total number of 15 acquisitions per bearing type was finally obtained. Each acquisition consists usually in 8 seconds of accelerations at a sampling frequency of 102.400 kHz. For each acquisition, also oil and bearing casing temperatures are measured using two thermocouples.

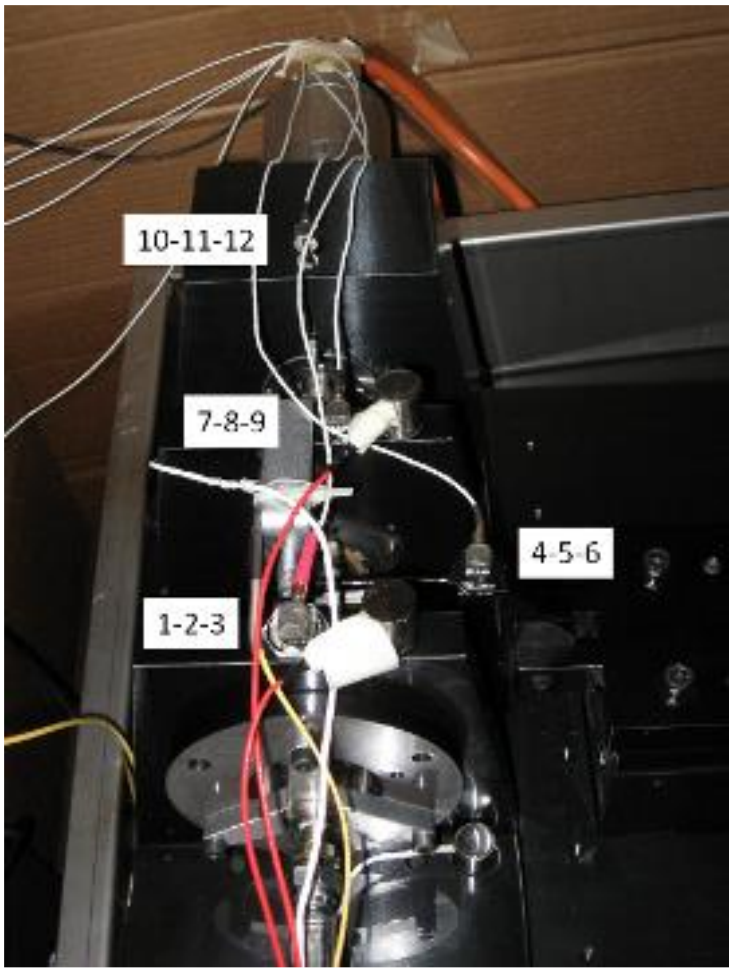


Figure 4.6- The setups used for the datasets acquisitions.

4.2 Test rig 2

The experimental setup detailed in Figs. 4.7 and 4.8 was used to record signals corresponding to the time-varying conditions. The test rig 2 assembled at the Dynamics of Advanced Structures and Machinery (DASM) Group at Department of Mechanical and Aerospace Engineering of Strathclyde University. It consists of a shunt DC motor (1 hp and 2000 rpm), bearing assembly, a mechanical loading system, an accelerometer and a portable data acquisition system.

4.2.1 Description

The data acquisition system used to collect acceleration signals is the portable *reactec* system shown in Fig. 4.9. Its frequency range is between 2 Hz and 40 kHz. The time block length which can be measured is 256 to 32768 samples.

The accelerometer used in the test rig 2 is MTN/1120 model and shown in Fig. 4.9. It is fixed to the bearing supports by magnetic coupling. The technical characteristics of the accelerometer is introduced in Table 4.3.

The bearings used in the experiment are SKF deep grooves 6308. The defect created for the faulty bearing used in the experiment was a very small notches on the inner race. Two bearing conditions are considered in this study (i.e. healthy and inner race faulty) and 20 signals are

acquired for each signal class. For all the acquired signals, the run up time varying operating condition are considered by varying the speed from 150 to 1500 rpm.

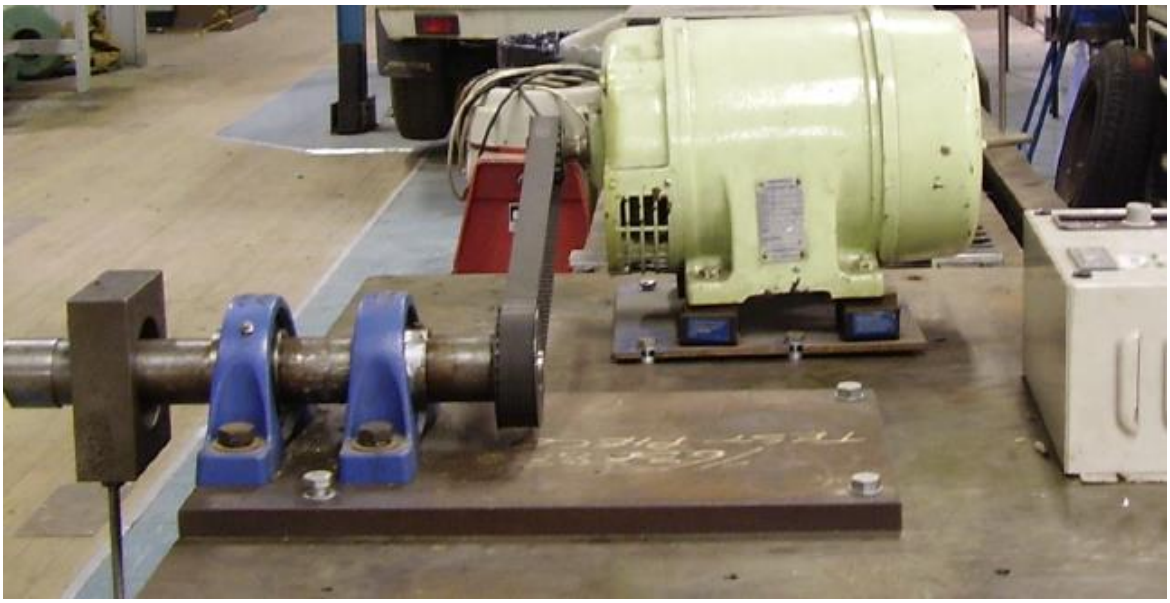


Figure 4.7- The experimental setup.

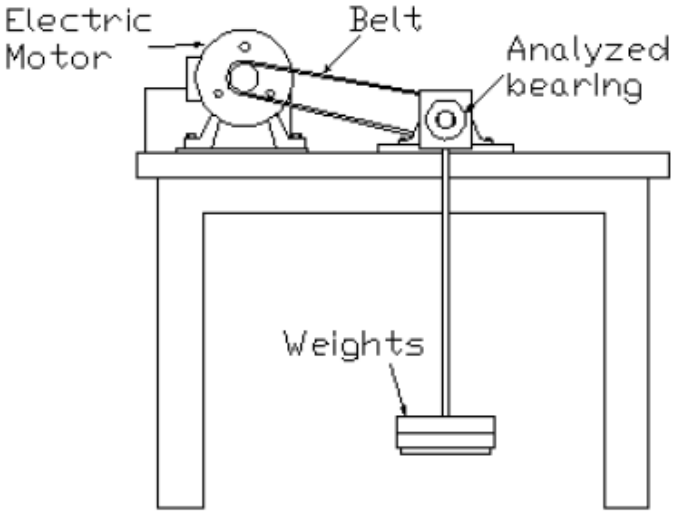


Figure 4.8- Sketch of the installation used.

Figure 4.9- The data acquisition system and accelerometer used in the test rig.

Table 4.3- Technical characteristics of the accelerometers

Producer	Model	Dynamic range	Nominal sensitivity	Mounted base Resonance
MONITRAN	MTN/1120	±80 g	100 mV/g	20 kHz

Chapter 5

Analysis and results

In this chapter, the Diagnostic techniques presented in Chapter 2 and 3 are applied to real data acquired on the test rig 1 and the test rig 2, described in chapter 4. As mentioned in section 2.2.1, one of the drawbacks of the EMD is sifting stop criterion, which might affect fault diagnosis results. In section 5.1, the criteria proposed for sifting stop of the EMD technique are investigated whether affect the damage detection of roller bearings. The signals (collected on the test rig1) are decomposed by the EMD using different stopping criteria and the feature vectors obtained from IMFs are classified using SVM to compare the labelling success rates and the calculated Margin.

Then, a combination method is proposed for early damage detection of roller bearing. The WPT technique is applied to data (collected on the test rig1) for denoising and the clean data is break-down into IMFs using the EEMD method.

In the section 5.3, in order to enhance the EEMD, a Performance improved EEMD (PIEEMD) is proposed to determine appropriate value for the amplitude of added noise.

A novel feature extraction method is proposed in the section 5.4 for identifying the small size defect based on the model created using only healthy bearing signals. The TKEO is applied to IMFs decomposed by PIEEMD technique to create new feature vectors as input data for one-class SVM.

Finally, a cointegration based method is proposed for fault diagnosis of roller bearing working in time-varying condition. The signals collected on test rig 2 are divided into some segments and each segment is decomposed by PIEEMD. The cointegrating vectors of first three IMFs are obtained by utilizing cointegration method. Then, TKEO is applied to each cointegrating vector to create three dimensional feature vectors used to construct the hyperplane of one-class SVM.

The results of this work have been published in some articles [252-256].

5.1 Sifting stop criterion of EMD

Although the EMD has been successfully used in various applications, there exist some drawbacks such as lack of a mathematical base, no robust stopping criterion for sifting process, mode mixing and border effect problem. One of the most relevant drawbacks in fault diagnosis is the sifting stop criterion. By definition, an IMF has to satisfy the condition whose mean between the upper and lower envelopes must be zero. Thus, the sifting process is repeatedly performed to make the symmetric zero mean modes. On the other hand, too many sifting steps will reduce the physical meaning of IMFs which are extremely important for fault diagnosis. Thus, an appropriate criterion is required to identify the step when the sifting process is concluded. The criteria proposed so far are: Cauchy-type convergence, Mean fluctuations thresholds, Energy difference tracking, Resolution factor, Bandwidths, and Orthogonality criterion, which are discussed in the following.

Cauchy-type convergence (SD)

Huang et al. [257] proposed a criterion where the size of the standard deviation-SD- of two consequent sifting results (h_n,h_{n-1}) should be limited and when it reaches a certain predefined value, sifting must stop:

$$SD = \sum_t \frac{[h_{n-1}(t) - h_n(t)]^2}{h^2{}_{n-1}(t)} \tag{5.1}$$

The main flaw of this approach is that it is unrelated to the definition of IMFs.

Mean fluctuations thresholds (MFT)

Rilling et al. [258] introduced a new criterion based on two thresholds (θ_1, θ_2) aimed at guaranteeing globally small fluctuations in the mean while taking into account locally large excursions. For $(1-\alpha)$ fraction of data, sifting will be continued when $\sigma(t) < \theta_1$ and for remaining fraction when $\sigma(t) < \theta_2$. The fraction $\sigma(t)$ is defined as follows:

$$\sigma(t) = \left|\frac{m(t)}{a(t)}\right| \tag{5.2}$$

where $a(t)$ and $m(t)$ are defined as:

$$a(t) = (e_{max}(t) - e_{min}(t))/2$$

$$m(t) = (e_{max}(t) + e_{min}(t))/2$$

$e_{max}(t)$ and $e_{min}(t)$ are the upper and lower envelopes, respectively.

They suggested a setting, in which $\alpha \approx 0.05$, $\theta_1 \approx 0.05$ and $\theta_2 \approx 10\,\theta_1$. The shortcoming of MFT is that the thresholds are not adaptive.

Energy difference tracking (EDT)

The energy difference tracking (EDT) is based on the assumption that the residue and IMFs are mutually orthogonal [259].

The total energy of the signal is computed as follows:

$$E_x = \int_{-\infty}^{+\infty} x^2(t)\, dt \tag{5.3}$$

Due to orthogonality of the components:

$$\int_{-\infty}^{+\infty} x_i(t)\, x_j(t)\, dt \cong 0 \tag{5.4}$$

where $i \neq j$. Thus, the total energy is represented as the sum of energy of the components:

$$E_x = \int_{-\infty}^{+\infty} x_1^2(t)\, dt + \int_{-\infty}^{+\infty} x_2^2(t)\, dt + \cdots + \int_{-\infty}^{+\infty} x_n^2(t)\, dt \tag{5.5}$$

The sum of energy of decomposed IMF (h_1) and those of the residual signal is formulated as follows:

$$E_{tot} = \int_{-\infty}^{+\infty} h_1^2(t)\, dt + \int_{-\infty}^{+\infty} [x(t) - h_1(t)]^2\, dt \tag{5.6}$$

If h_1 is an orthogonal component of $x(t)$, the E_{tot} equals to the original signal energy E_x. Otherwise, there exists a difference denoted as E_{err}. Hence, E_{err} is tracked as the stopping criterion and when it reaches a certain minimum and the mean value of envelope becomes small enough, sifting process is terminated.

$$E_{err} = E_{tot} - E_x = \int h_1^2(t)dt - \int x(t)\, h_1(t)\, dt \tag{5.7}$$

Resolution factor (RF)

Rato et al. [260] applied the ratio between the energy of the signal at the beginning of the sifting (E_x) and the energy of mean of envelopes (E_m) as the sifting stop criterion. The process is terminated when the ratio ascends a predefined factor (resolution factor (RF)).

$$\frac{E_x}{E_m} \leq RF \tag{5.8}$$

Setting the RF criterion is determined by practice for each signal and they set it 40 dB to decompose an Electro-Cardio-Gram (ECG) signal and 50 dB for an Electro-Encephalo-Graphy (EEG) signal.

Bandwidth (BW)

Xuan and Xie [261] designed a new stop criterion based on two types of bandwidth: instantaneous bandwidth and frequency bandwidth which is caused only by frequency changes. It uses the MFT criterion to get a result that almost satisfies the two conditions of IMF. Then, sifting process is continued until to find the minimum of σ_f^2 (variance of instantaneous frequency) or until the difference of σ_f^2 between two consequent sifting results is very small:

$$\left|(\sigma_f^2)_{h_n} - (\sigma_f^2)_{h_{n-1}}\right| < \varepsilon \tag{5.9}$$

Orthogonality criterion (OC)

Based on orthogonality definition, Lin and Hongbing defined orthogonality criterion (OC) [262]:

$$OC = \left| \sum_{t=1}^{N} \frac{m(t)\, x(t)}{m(t)\,\big(x(t) - m(t)\big)} \right| \tag{5.10}$$

when OC reaches to a pre-defined value, the sifting will be terminated.

If $m(t) = b\, x(t)$ so that $0 < b < 1$, then $OC = \frac{1}{1-b} > 1$. If $m(t)$ is irrelative to $x(t)$, then $OC < 1$. They set OC > 1.05 as the sifting condition in the inverse EMD filter. They applied it to a vibration signal to show that their improved EMD, which includes some modifications such as the proposed stopping criterion, achieves less mode mixing problem.

Methodology

The goal is to evaluate the criteria to identify which stopping criterion is more efficient for bearing damage detection. The bearing data set (acceleration signals) were collected under various operating conditions and different fault severities using the test rig 1 (section 4.1) developed and assembled by the Dynamics & Identification Research Group (DIRG) at the Department of Mechanical and Aerospace Engineering of Politecnico di Torino. The data collected with three different shaft speeds (200, 300 and 400 Hz) and three different external radial loads (1.0, 1.4 and 1.8 kN) are utilized. The original acquired signals are divided into 10 segments including 10000 data points each, to extract required informative feature vectors. Thus, each signal includes 10 segments.

The fault diagnosis method is given as the following:

1) To apply EMD with different sifting stop criteria to decompose each segments into some IMFs. The first m IMFs which include the most dominant fault information are chosen to extract the feature.
2) To calculate the total energy E_i of the first m IMFs:

$$E_i = \int_{-\infty}^{+\infty} |c_i(t)|^2\, dt \tag{5.11}$$

3) To create a feature vector with the energies of the m selected IMFs:

$$FV = [E_1, E_2, \dots, E_m] \tag{5.12}$$

4) To normalize the feature:

$$FV_n = [E_1/E\,, E_2/E\,, \dots, E_m/E] \tag{5.13}$$

where $E = (\sum_{i=1}^{m} |E_i|^2)^{1/2}$

5) To construct the SVM by utilizing the normalized feature vectors. The 60% of data are used for training and the rest are taken as the test samples.

Results and analysis

By implementing the methodology described, feature vectors for each sifting stop criterion and defect size are obtained. The IMFs obtained using the SD criterion are presented in Fig. 5.1. The normalized energy of IMFs introduced as an efficient feature vector in fault diagnosis

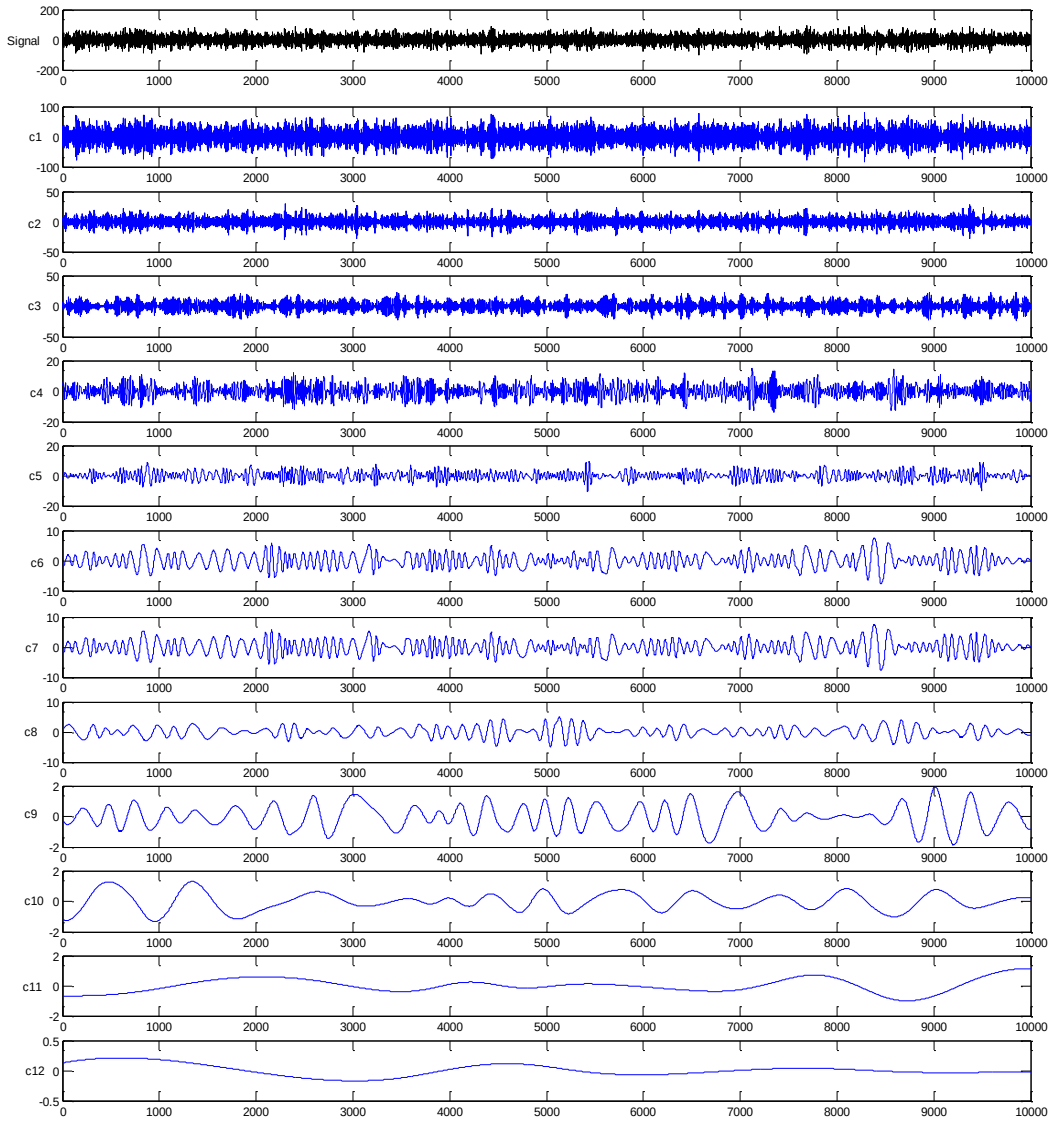
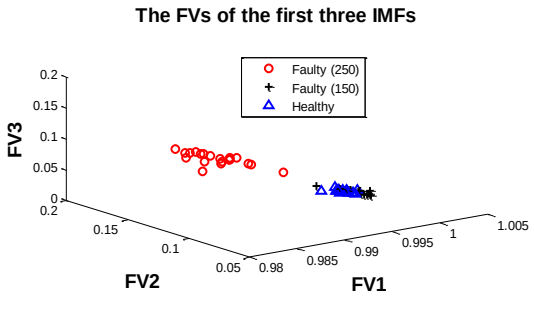


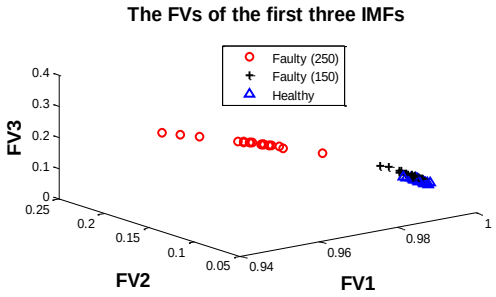
Figure 5.1- A real collected vibration signal of a roller bearing and the obtained IMFs using the EMD method

of roller bearing, has been adopted using only the first three elements of the feature vectors. In Fig. 5.2, feature vectors for healthy and defective bearings (150 and 250 microns defect levels) are shown for a specific operating condition (Shaft speed = 200 Hz and external load = 1.4 kN). For bigger size damage (250 microns), two classes of data (healthy and damaged) are perfectly separated for all criteria. Whereas, for smaller defect (150 microns), it seems that still exists a confusion between the two classes in some cases such as the SD algorithm. For example, the results in Figs. 5.3 and Fig. 5.4, show the data confusion adopting the SD criterion, whereas results are distinctly divided into two groups applying the MFT. The classification results of the samples obtained by applying all sifting stop criteria and for various operating conditions

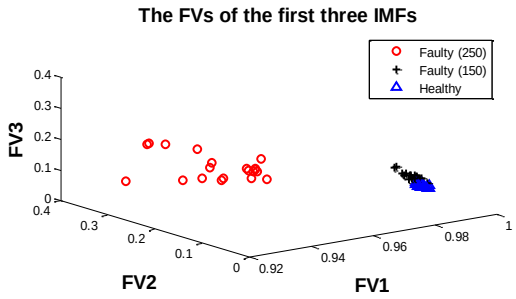
are proposed in Table 5.1 and Table 5.2 for small and big defects, respectively. For those cases where the test success rate is perfect (100%), the calculated margin is presented, instead. Comparing the success rates proposed in Tables 5.1 and 5.2, it can be found that the most accurate defect detection can be achieved by using MFT, EDT and RF criteria. In fact only in two cases the detection rates are not perfect for these criteria (considering both damage sizes). It seems that the worst result of the classification rates have been acquired applying the SD criterion so that in eight states there exist some wrong labeled samples.



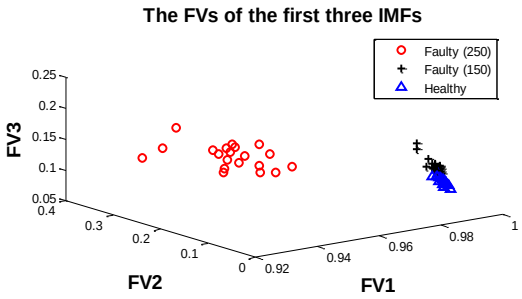
a) SD criterion



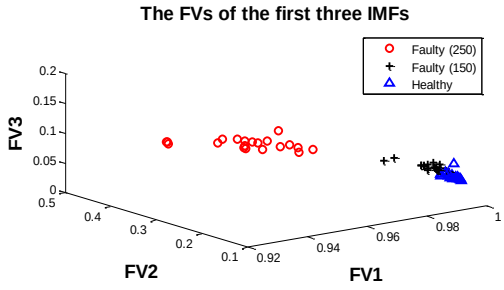
b) MFT criterion



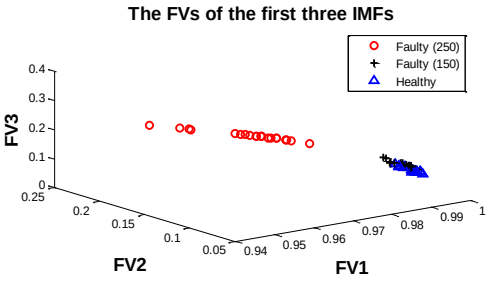
c) EDT criterion



d) RF criterion



e) BW criterion



f) OC criterion

Figure 5.2- 3D Feature vectors of all criteria at speed = 200 Hz and load = 1.4 kN

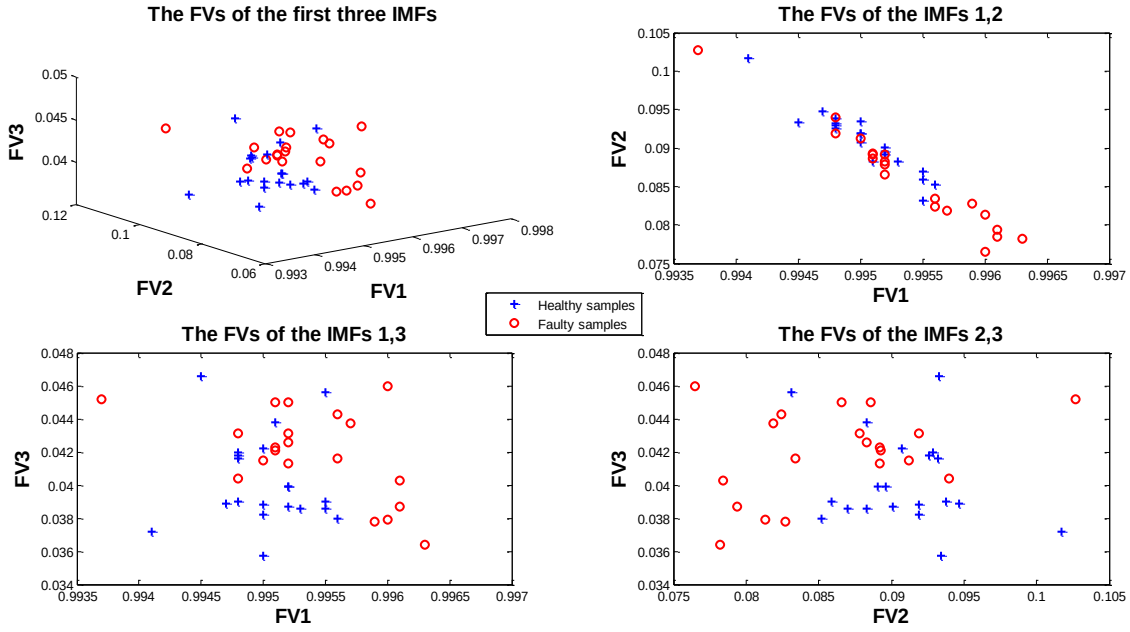


Figure 5.3- Feature vectors of SD criterion at speed = 200 Hz and load = 1.4 kN

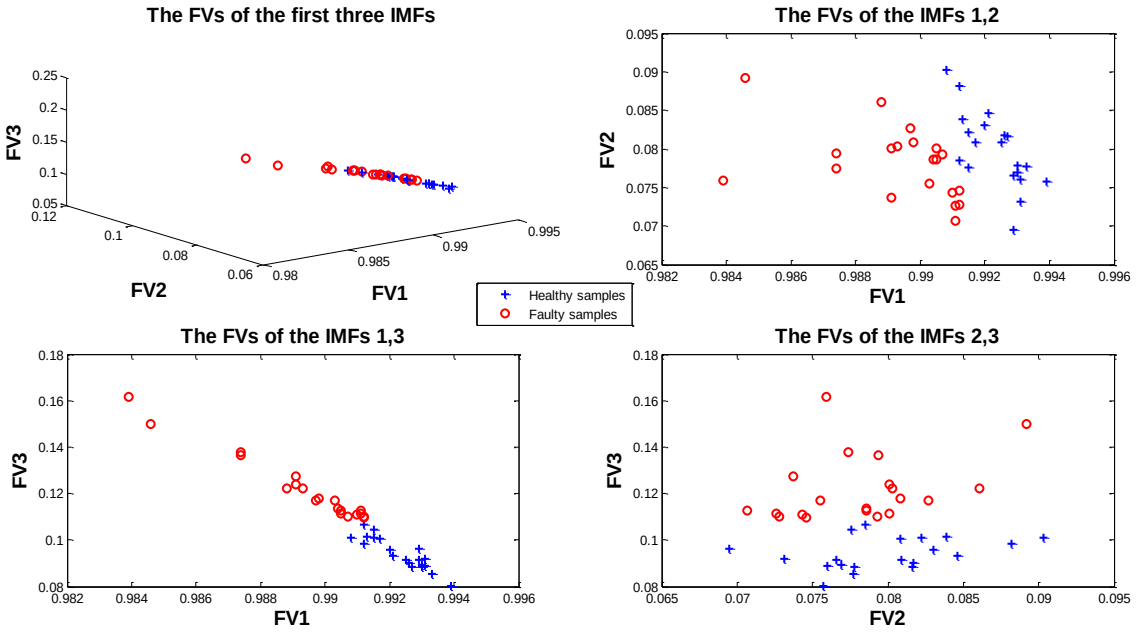


Figure 5.4- Feature vectors of MFT criterion at speed = 200 Hz and load = 1.4 kN

For those conditions who has achieved perfect classification rates, Margin (whose meaning was explained in section 3.3) is considered in order to compare them. It does not exist a clear trend for margin to determine which one creates more reliable classification structure. However, higher mean values of the Margins has been obtained by EDT criterion for both defect severities.

As it can be seen in Tables 5.1 and 2.2, the accuracy of fault diagnosis tends to decrease where higher external loads are applied. In the case of higher level of defect, misclassification occurs at

higher shaft speeds whereas for small level, such an incorrectly labelling exists even for lower speed.

In next step, all the data which belongs to different speeds and loads are considered in a whole, to investigate how efficient each method is, to remove the effects of speed and external load. Nine operating conditions (three speeds and three loads) and three bearing conditions (normal, small and big defect sizes on Roller (150 and 250 microns) are investigated; 180 samples (9×2×10×1) for each criterion are divided into 108 training and 72 test samples. The defect identification accuracy of different methods is shown for both damage levels in Table 5.3. As it can be seen, the most accurate results for smaller defect level is achieved using EDT criterion (93.5% classification rate for training and 80.6% for test) and for higher damage level by using RF criterion (99.1% classification rate for training and 97.2% for test). The lowest success rates for both levels is obtained using SD criterion.

Table 5.1- Margin (In the case of 100% success rate) and success rate -in bold- of fault diagnosis (Defect size 150 microns)

Criterion	200 (Hz) 1.0 (kN)	200 (Hz) 1.4 (kN)	200 (Hz) 1.8 (kN)	300 (Hz) 1.0 (kN)	300 (Hz) 1.4 (kN)	300 (Hz) 1.8 (kN	400 (Hz) 1.0 (kN)	400 (Hz) 1.4 (kN)	400 (Hz) 1.8 (kN)
	Margin	Margin	Margin	Margin	Margin	Margin	Margin	Margin	Margin
SD	1.3170	**87.5%**	**87.5%**	1.0695	**87.5%**	**87.5%**	**75.0%**	1.2141	1.1310
MFT	1.3609	0.9137	1.0093	1.1867	**87.5%**	0.9849	1.0178	0.9556	1.0041
EDT	1.3224	1.0078	**87.5%**	1.2530	1.2973	1.1433	**87.5%**	1.1752	1.1623
RF	1.3397	**87.5%**	1.0940	1.3213	1.0109	0.9877	1.0198	1.2151	1.0738
BW	1.3738	1.1292	**87.5%**	1.3784	1.1359	1.3151	1.0427	1.1069	1.1265
OC	1.3261	0.9798	0.9733	1.2131	**87.5%**	**87.5%**	1.0901	1.0301	1.0893

Table 5.2- Margin (In the case of 100% success rate) and success rate-in bold- of fault diagnosis (Defect size 250 microns)

Criterion	200 (Hz) 1.0 (kN)	200 (Hz) 1.4 (kN)	200 (Hz) 1.8 (kN)	300 (Hz) 1.0 (kN)	300 (Hz) 1.4 (kN)	300 (Hz) 1.8 (kN)	400 (Hz) 1.0 (kN)	400 (Hz) 1.4 (kN)	400 (Hz) 1.8 (kN)
	Margin	Margin	Margin	Margin	Margin	Margin	Margin	Margin	Margin
SD	1.2922	1.3947	1.1601	1.2121	**87.5%**	**87.5%**	1.0479	0.9910	**75.0%**
MFT	1.3725	1.3769	1.3275	0.9529	1.0497	**87.5%**	1.1075	1.1053	1.1013
EDT	1.4079	1.3347	1.1917	1.2991	1.3496	1.3565	1.2393	1.2559	1.2580
RF	1.3913	1.3729	1.3859	1.0668	1.0882	**87.5%**	1.1029	1.0508	1.2163
BW	1.3884	1.3867	1.3060	1.0622	1.1507	0.8300	1.1005	**87.5%**	**75.0%**
OC	1.3543	1.3453	1.3341	**75.0%**	1.0637	1.0006	1.3578	1.3360	1.2039

Table 5.3- Success rate of speed and load effects removal and fraction of correctly labeled samples for both defect sizes

Criterion	150 microns				250 microns			
	Training		Test		Training		Test	
	Success rate	Fraction of correctly labeled	Success rate	Fraction of correctly labeled	Success rate	Fraction of correctly labeled	Success rate	Fraction of correctly labeled
SD	73.1%	(79/108)	72.2%	(52/72)	87.0%	(94/108)	83.3%	(60/72)
MFT	84.3%	(91/108)	83.3%	(60/72)	97.2%	(105/108)	95.8%	(69/72)
EDT	93.5%	(101/108)	80.6%	(58/72)	96.3%	(104/108)	94.4%	(68/72)
RF	85.2%	(92/108)	84.8%	(61/72)	99.1%	(107/108)	97.2%	(70/72)
BW	88.0%	(95/108)	83.3%	(60/72)	93.5%	(101/108)	91.7%	(66/72)
OC	82.4%	(89/108)	81.9%	(59/72)	95.4%	(103/108)	93.1%	(67/72)

Conclusions

It has been shown that the highest success rate in the recognition process of the bearing state, is obtained using MFT, EDT and RF criteria. Although higher mean value of margin is achieved using EDT, the mean values obtained by MFT and RF are not far less. In addition, in order to remove speed and load effects contemporary, the most accurate results are obtained by using EDT and RF criteria for small defect size and higher level of defect, respectively.

5.2 Early damage detection using EEMD and denoising

Acceleration signals acquired are usually noisy, which significantly affects the results of fault diagnosis. Thus, the signals must be denoised before applying diagnostic techniques. The denoising methods were discussed in section 2.1. Wavelet packet Transform (WPT) is a powerful method utilized effectively for the denoising procedure of the acquired signals. Furthermore, based on the results of section 5.1, in some cases, small defect size are not perfectly detected using the EMD technique. Therefore, a more powerful diagnostic method has been developed. As mentioned in section 2.2.2, the EEMD method has been proposed to eliminate one of the drawbacks, i.e. the mode mixing problem of the EMD technique. Two important parameters which may influence the performance of the EEMD are the amplitude of added noise and the number of ensemble trials. If the amplitude of the added noise is too small with respect to the original signal, a considerable mode mixing improvement cannot be achieved. On the other hand, if the amplitude of the added noise is too large, it will create some redundant IMF components which lead to misinterpretation of the analysis result. Although an infinite number of ensemble trials is needed to completely cancel out the effect of the added white noise, too many trial numbers would increase the computational cost. Wu and Huang [171] suggested 0.2 of standard deviation of the original signal for the amplitude of the added white noise and a few hundred ensemble trials. However, it seems that there is no robust guide to choosing the optimum amplitude from the wide range suggested.

In this section a combined method is investigated to detect very small defects on roller bearings. The WPT is applied to clean the noisy signals, then, informative feature vectors are extracted using the EEMD technique.

Methodology

It is investigated whether the proposed combined method is able to detect very small defects on roller bearings. The bearing data set (acceleration signals) were collected under various operating conditions and different fault severities using the test rig 1 (section 4.1). Three different shaft speeds (200, 300 and 400 Hz) and three different external radial loads (1.0, 1.4 and 1.8 kN) are considered to acquire the signals in different operating conditions. The early defect over one roller is 150 microns in diameter. The original acquired signals were divided into 20 segments including 10000 data points each, to extract required informative feature vectors. The Normalized energy of IMFs are used as the feature vectors.

The fault diagnosis method is given as the following:

1) To apply WPT to extract the clean data from the collected noisy signals.
2) To apply EEMD with different amplitude of added white noise (from 0.2 to 0.6 of standard deviation of the original signal) to decompose the vibration signals into some IMFs. The first m IMFs including the most dominant fault information are chosen to extract the feature.
3) To calculate the Normalized feature vectors as explained in section 5.1.
4) To construct the SVM by utilizing the normalized feature vectors. 60% of data are used for training and the rest are taken as the test samples.

Results and analysis

By implementing this methodology the normalized energy of the IMFs is calculated using only the first three IMFs. In Table 5.4, it is shown that the small defect size (150 microns) is not recognized using the EMD for some working conditions. Increasing speed and load lead to less success rates. From Fig. 5.5, it is clear that healthy and faulty samples cannot be separated completely and consequently the labeling success rate will not be perfect.

The present step is hence devoted to investigate whether the proposed combined method is able to better detect the defect for those working conditions in which the EMD was not able to do so (Table 5.4). In order to eliminate the riding noise from the noisy signal, the WPT is applied using the discrete Meyer and Daubechies mother wavelet. In Fig. 5.6, the noisy signal, the separated noise and the denoised signal are shown for a specific working condition (Speed = 300 Hz and load = 1.4 kN). The signal to noise ratios (SNR) presented in Table 5.5 (for both healthy and faulty signals) reveals that the use of WPT allows to clean the data effectively, especially for the lower speed. For those load and speed settings whose damage detection was unsatisfactory by simply using EMD (Table 5.4), it is now possible to identify the defect adopting the proposed combined techniques (Table 5.6).

Table 5.4- Success rates of classification for different operating conditions using EMD

EMD	200 (Hz) 1.0 (kN)	200 (Hz) 1.4 (kN)	200 (Hz) 1.8 (kN)	300 (Hz) 1.0 (kN)	300 (Hz) 1.4 (kN)	300 (Hz) 1.8 (kN)	400 (Hz) 1.0 (KN)	400 (Hz) 1.4 (kN)	400 (Hz) 1.8 (kN)
Training	100	100	100	100	100	100	100	100	100
Test	100	100	**81.3**	100	**81.3**	100	100	**93.8**	**62.5**

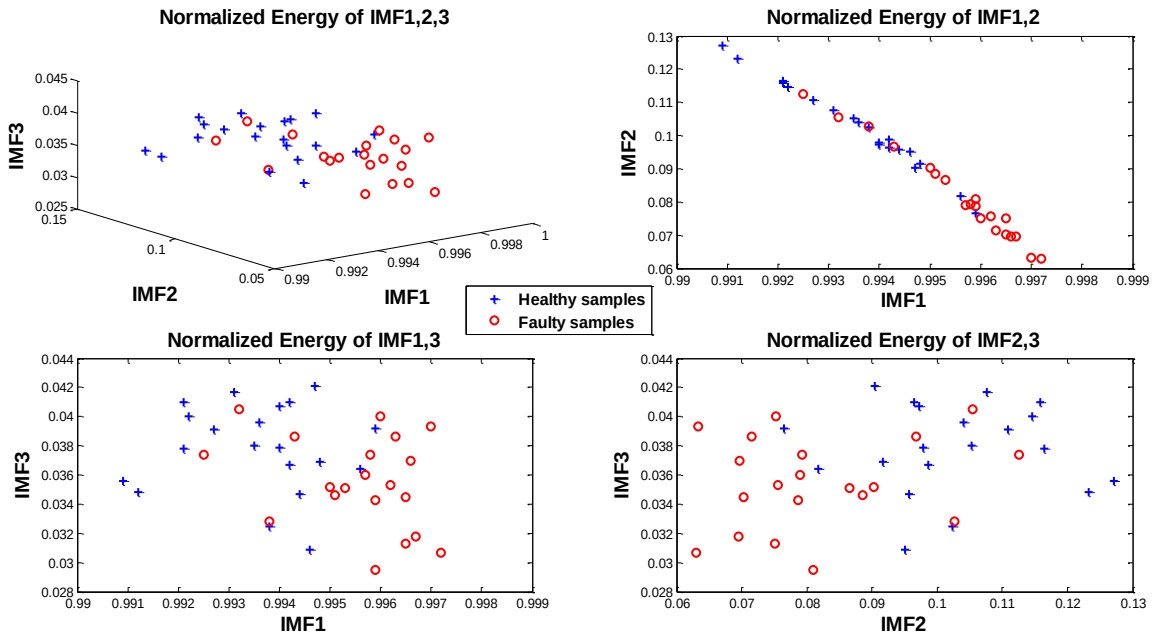
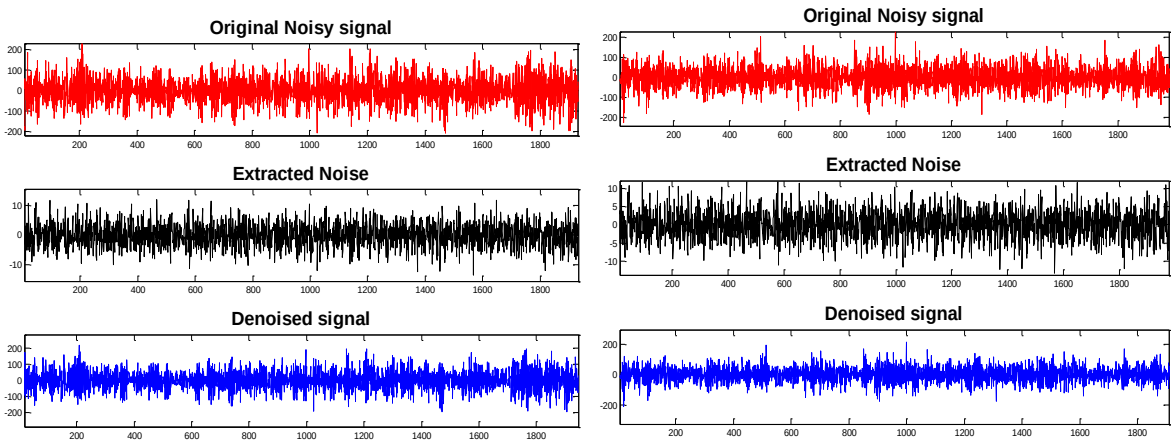


Figure 5.5- Normalized energy of first three IMFs for the worst condition (Speed = 400 Hz and load = 1.8 kN)



(a) (b)

Figure 5.6- Denoising signals using db10 (Speed = 300 Hz and load = 1.4 kN). a) Healthy signal b) Faulty signal

Table 5.5- SNR ratio for healthy and faulty bearing signals of each operating conditions (After applying WPT)

WPD	200 (Hz) 1.8 (kN)		300 (Hz) 1.4 (kN)		400 (Hz) 1.4 (kN)		400 (Hz) 1.8 (kN)	
	Healthy	Faulty	Healthy	Faulty	Healthy	Faulty	Healthy	Faulty
SNR	12.0	12.4	22.8	21.3	27.8	27.3	26.4	27.2

Table 5.6- Fault diagnosis success rate for those operating conditions whose state was not possible to be recognized using EMD algorithm

Method	Added noise amplitude	200 (Hz) 1.8 (kN)		300 (Hz) 1.4 (kN)		400 (Hz) 1.4 (kN)		400 (Hz) 1.8 (kN)	
		Training	Test	Training	Test	Training	Test	Training	Test
EMD	-	100	81.3	100	81.3	100	93.8	100	62.5
EMD_Meyer	-	100	87.5	100	87.5	100	93.8	100	62.5
EMD_ db10	-	100	93.8	100	87.5	100	93.8	100	75.0
EEMD	0.2	100	87.5	100	87.5	100	100	100	100
EEMD	0.3	100	87.5	100	93.8	100	100	100	100
EEMD	0.4	100	81.3	100	100	100	100	100	100
EEMD	0.5	100	87.5	100	100	100	100	100	100
EEMD	0.6	100	87.5	100	100	100	100	100	100
EEMD_Meyer	0.2	100	93.8	100	87.5	100	100	100	100
EEMD_Meyer	0.3	100	93.8	100	87.5	100	100	100	100
EEMD_Meyer	0.4	100	93.8	100	100	100	100	100	100
EEMD_Meyer	0.5	100	93.8	100	100	100	100	100	100
EEMD_Meyer	0.6	100	93.8	100	93.8	100	100	100	100
EEMD_db10	0.2	100	100	100	87.5	100	100	100	100
EEMD_db10	0.3	100	100	100	87.5	100	100	100	100
EEMD_db10	0.4	100	93.8	100	93.8	100	100	100	100
EEMD_db10	0.5	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
EEMD_db10	0.6	100	93.8	100	100	100	100	100	100

Some interesting results are:

- Although denoising improves the results in many cases, a perfect classification accuracy is not obtained using denoising in a combination with the EMD technique. Denoising with the Daubechies (db10) mother wavelet achieves higher success rates than the Meyer mother wavelet.
- By applying the EEMD to the noisy signals (amplitude of 0.4, 0.5 and 0.6 of standard deviation of the original signal), in the most cases the condition is perfectly recognized (except for the condition with shaft speed of 200Hz and 1.8kN load).
- For the noisy signals, the most appropriate amplitude of added noise seems to be 0.5 or 0.6 of standard deviation of the original signal.

Applying the EEMD (amplitude of 0.4 or 0.5) to the signals denoised by the Meyer mother wavelet, the state of bearing in the most cases is correctly recognized. It improves the result for the previously mentioned condition (speed of shaft = 200 Hz and the load = 1.8 kN).A perfect success rates is achieved by denoising with Daubechies mother wavelet (db10) and using the amplitude of 0.5 of standard deviation of the original signal for the all operating conditions. As an

illustrative instance (Fig. 5.7), healthy and faulty samples of the worst condition (using EMD (Fig. 5.5)) are completely separable.

For those operating conditions whose states were perfectly detected by using EMD (Table 5.4), only for one specific condition (Speed = 200 Hz and load = 1.4 kN) EEMD could not successfully identify the state. The success rates achieved applying the EEMD are 93.8%, 100%, 93.8%, 87.5% and 87.5% utilizing 0.2, 0.3, 0.4, 0.5 and 0.6 as the amplitude, respectively. Denoising improves the results and then EEMD can achieve the perfect classification rate for this specific condition.

Table 5.7- Defect detection success rate for speed and load effects removal

	Added noise amplitude	Training	Test
EMD	-	87.0	70.1
EMD_db10	-	88.0	69.4
EEMD	0.3	91.2	77.1
EEMD	0.5	90.3	75.7
EEMD	0.6	88.4	70.8
EEMD_db10	0.3	91.7	77.8
EEMD_db10	0.5	89.8	70.1
EEMD_db10	0.6	88.9	73.4

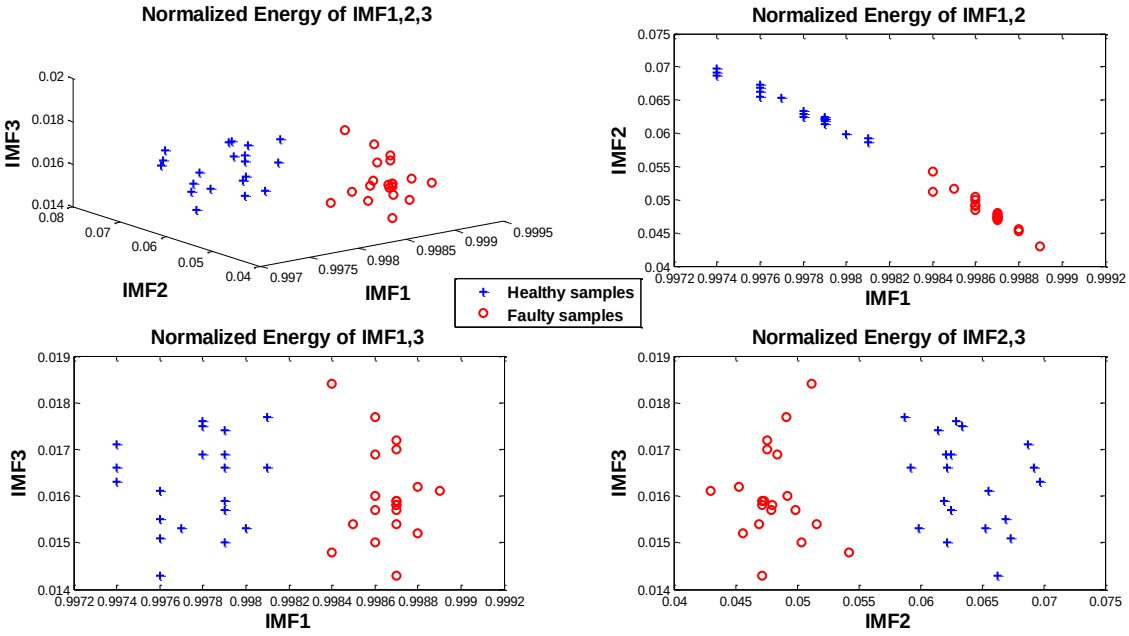


Figure 5.7- Normalized energy of IMFs using WPT (db10) and EEMD with noise amplitude = 0.5 standard deviation of the signal (Speed = 400 Hz and load = 1.8 kN)

Next, we consider all the data belonging to different speeds and loads together (at the same time) to investigate the performance of each method in removing the effects of speed and external load. The success rates of data labelling of different methods is shown in Table 5.7. As it can be seen, denoising could not improve the results, whereas using EEMD instead of EMD is increased the classification accuracy. The most accurate results is achieved by applying WPD (using db10 as the mother wavelet) and EEMD with 0.3 of the standard deviation of the signal as the amplitude of the added noise.

Thus, the most appropriate amplitude to obtain the perfect classification rate for each specific operating condition (with denoising by db10) are 0.5 and 0.6, whereas 0.3 seems to be the best for removing both speed and load effects.

Conclusions

In the denoising process, more accurate classification results are obtained using the Daubechies mother wavelet (db10). In addition, it is shown that the accuracy of damage identification is significantly affected by the amplitude of the added white noise (in the EEMD processing step). The new samples of all operating conditions are labeled perfectly by using db10 as a mother wavelet and 0.5 of standard deviation of the signal as the amplitude of the added noise. However, to remove speed and load effects contemporarily, the most accurate results are obtained by using smaller amplitude (0.3).

It is shown that the amplitude of added noise affects the results of early damage detection significantly. However, there is no robust guide to selecting an optimal amplitude for the added noise. In the next section, instead of using previously suggested amplitudes which are predefined and constant, a reliable and adaptive method will be investigated to determine an appropriate amplitude.

5.3 Performance improved EEMD (PIEEMD)

As discussed in section 5.2, the EEMD is a powerful technique for damage detection. However, to successfully handle this technique still remains a great challenge: how to better identify the two effective parameters (the amplitude of added noise and the number of ensemble trials), which affect the performance of the EEMD. If the amplitude of the added noise is too small relative to the original signal, a considerable mode mixing improvement cannot be achieved. On the other hand, if the amplitude of the added noise is too high, it will create some redundant IMF components which lead to misinterpretation of the analysis result. In addition, although, an infinite number of ensemble trials is required to completely cancel out the effect of the added white noise, too many trial numbers would increase the computational cost. Although a number of algorithms or values have been proposed, there is no robust guide to select optimal amplitude and the ensemble trial number yet, especially for early damage detection.

Wu and Huang [171] suggested the value of 0.2 of standard deviation of the original signal as the amplitude of the added white noise and a few hundred for trial number of ensemble. It has been shown in various cases that such an amplitude is not appropriate. Zhang et al. [263] suggested using a band-limited white noise to decrease the computational cost. Analyzing a simulated signal, it was concluded that appropriate range of SNR (signal to noise ratio based on signal power) is [50-60] dB. However, they used another range ([0.01-0.1]), which is outside of the suggested SNR. A non-stationary signal was constructed to mimics realistic vibration signals measured from rolling bearing and the appropriate range of SNR was considered [49-58] dB for the vibration signals. Applying the EEMD to the simulated signals, it was obtained that when the number of

ensemble trials is 100, the corresponding correlation coefficient approaches 0.95. Using the modified EEMD method, the acceptable results were achieved approximately after 70 ensemble trials, instead of the100 trials suggested for the original EEMD method. For real data (acceleration signals), it was shown that the percentage improvement of the computational efficiency (the consumed time ratio) varies from 30% to 45%, depending on the operating conditions. Guo and Tse [264] investigated the influence of the parameters setting on the results of reducing the mode mixing problem using a simulated signal. The effects of frequency and amplitude ratio of two different parts of the simulated signal (the high frequency and low frequency components) were investigated as well. The investigated amplitudes were considering again coefficients of standard deviation of the original signal (0.01, 0.1, 0.2, and 0.3). As real data is noisy (produced by other industrial equipment) and the amplitudes and composition of frequency are unknown, lower amplitude of noise was added and more number of ensemble trials applied (0.1 of standard deviation of the original signal for amplitude and 3000 for ensemble trial number). As only one specific operating condition with a single pre-defined amplitude was investigated, it would not represent a reliable guideline for properly setting the best parameters for real signals. Lin [265] tried to provide a guidance on choosing the appropriate amplitude and reduce the tremendous time waste occurring in the EEMD method. An optimal interval was suggested that lies between the square root of the average power of the weak sinusoid component and that of the weak transient component. When the amplitude is selected from the mentioned interval, the Pearson's correlation coefficients of the components reach their maximum value. Taking into consideration that only one specific gearbox vibration signal was investigated to verify the suggested procedure, its performance does not seem too reliable to identify small defects. Furthermore, applying such a procedure results difficult in damage identification, especially for automatic damage detection. Jiang et al. [266] applied multiwavelet packet as a pre-filter to enhance the weak multi-fault features in the narrow frequency bands. Then two ranges were suggested for the amplitude: [0-0.2] of the standard deviation of the original signal for high frequency components and [0.2-0.6] of the standard deviation of the original signal for the low frequency components. As some specific amplitudes were selected (0.04, 0.08 and 0.5) without any justification in this study, it seems that no robust guide is yet available to choose the optimum amplitude based on the wide suggested ranges. In section 5.2, It was concluded that more appropriate amplitude was [0.4-0.6] of the standard deviation for noisy signals, and 0.5 for denoised signals. The number of trial was set on 100 for all conditions. Furthermore, as mentioned in section 5.2, there is no suggestion for the ensemble trial number; it is only declared in three before-mentioned studies - the modified new number for simulated signals (70 trials) in [263], a very high number (3000 trials) in [264] and the 100 trials in [171,252].

Now, a reliable method is proposed to determine the suitable amplitude and the proper number of trials is investigated as well.

Performance improved EEMD (PIEEMD)

Here, an adaptive method, Performance Improved EEMD (PIEEMD) is proposed and its performance and applicability are evaluated utilizing several real vibration signals. After adding a random white noise, by applying the SNR definition (Eq. 5.14), the Amplitude value for each data point of a sample is obtained from Eq. 5.15. Considering an appropriate value for SNR, there would be a confidence that the extrema of the original signal are influenced adequately.

$$SNR(t) = 20\ \log\big(x(t)/(Amp_j(t)\cdot\ n_j(t))\big)$$

(5.14)

$$Amp_j(t) = 10^{-(SNR(t)/20)}\ \cdot (x(t)/n_j(t))$$

(5.15)

where $j = 1, \dots, M$ (j is the is the ensemble trial number). Then, the noise added signal is formulated as follows (as explained in section 2.2.3):

$$x_j(t) = x(t) + Amp_j(t)\cdot n_j(t)$$

(5.16)

In Fig. 5.8, a vibration signal of a roller bearing and a created random noise are shown. A suggested fixed value (0.3) multiplied by standard deviation of the original signal, creates a predefined constant value along the whole signal (Fig. 5.9). Thus, affecting on the extrema depends on value of random noise at the location of the extrema. Using the proposed algorithm (Eq. 5.15), an adaptive value (Fig. 5.9) is generated to preserve the SNR ratio. It means that for any randomly created noise, the amplitude will be high enough to affect the extrema. Investigating the result of adding noise to the vibration signal shows how the proposed amplitude acts more efficiently on the extrema (Fig. 5.10).

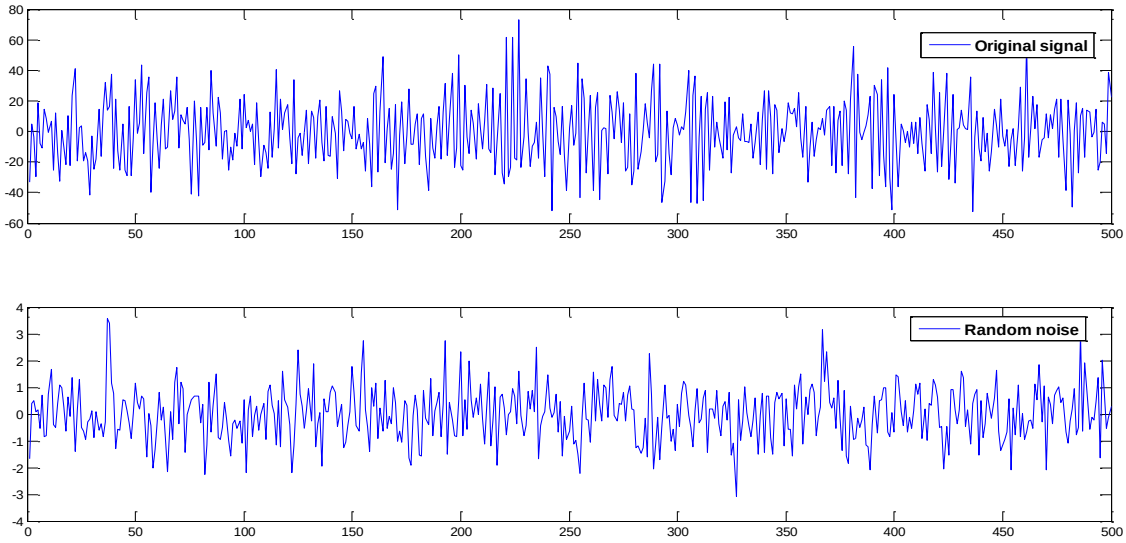


Figure 5.8- A real vibration signal of a roller bearing (up) and a random white noise (down)

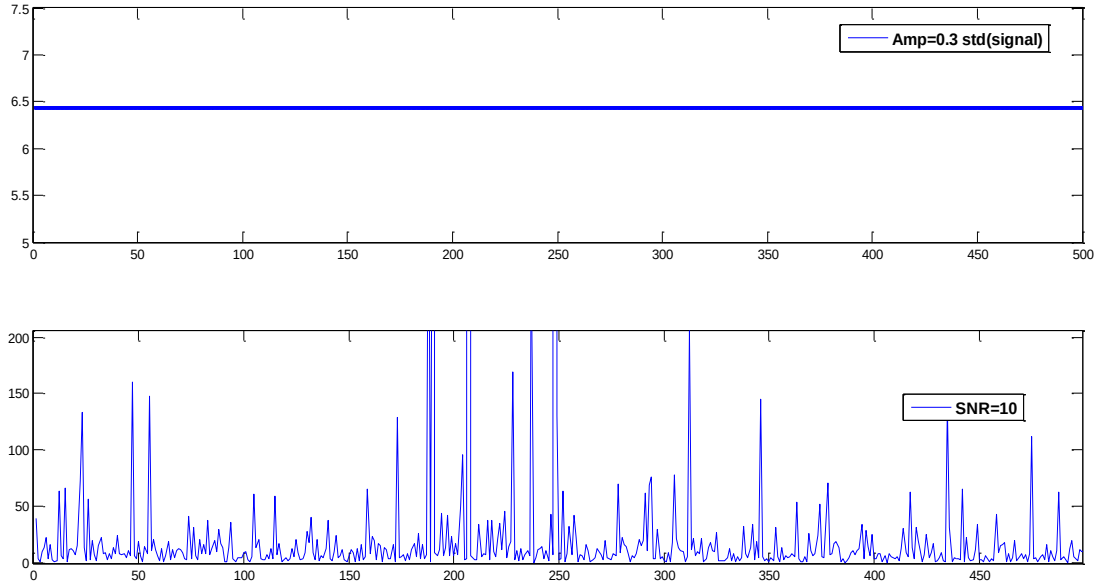


Figure 5.9- The predefined constant amplitude (0.3 standard deviation of the signal)) (up) and the proposed amplitude algorithm (PIEEMD with SNR = 10) (down)

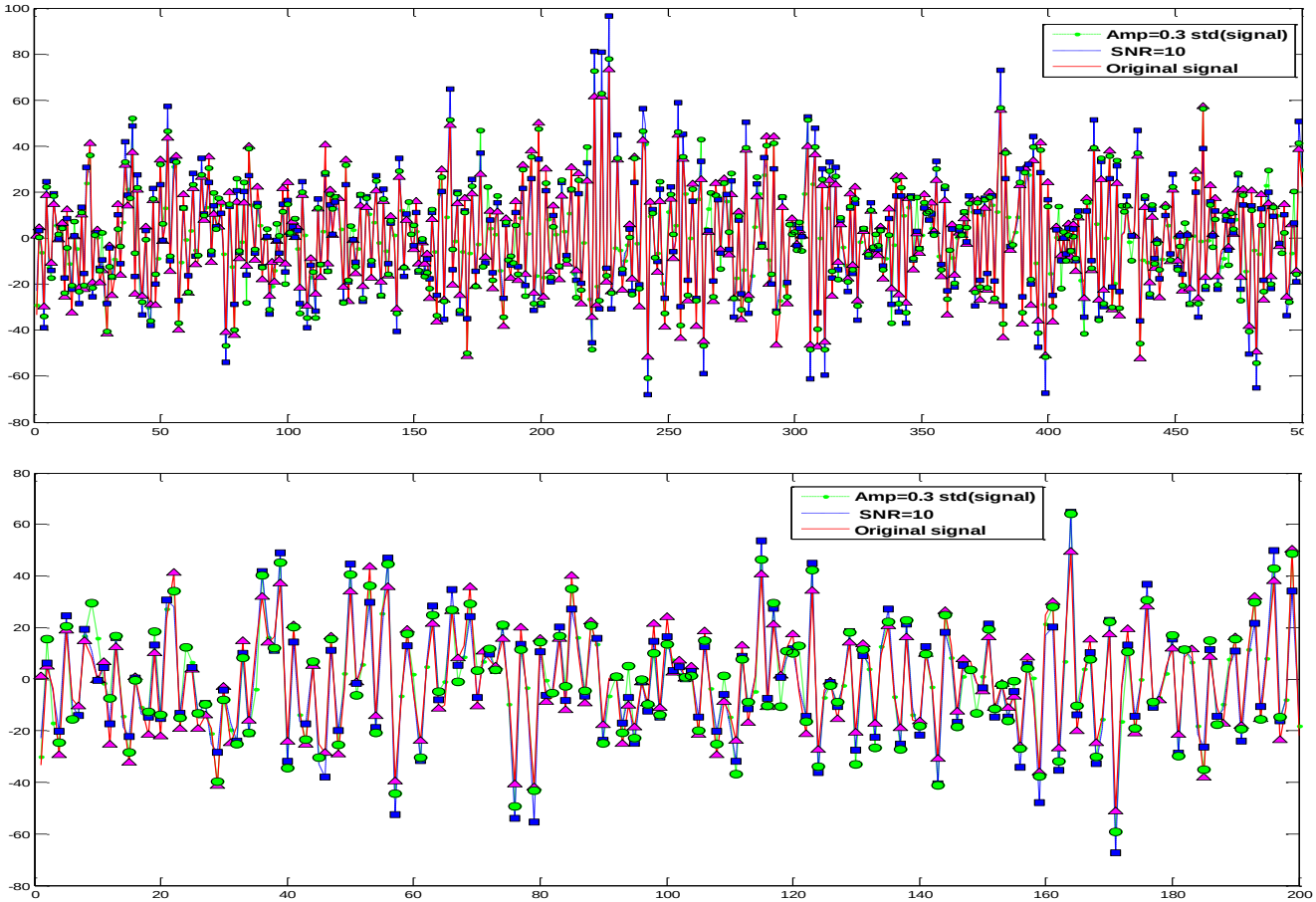


Figure 5.10- Influence on extrema, using constant amplitude (0.3 standard deviation of the signal)) and the PIEEMD algorithm (SNR = 10)

Methodology

The goal is to evaluate the performance of the proposed amplitude calculation algorithm, performance improved EEMD (PIEEMD), for the various operating conditions of a roller bearing. The fault diagnosis method is as follows:

1) To apply the EEMD with different amplitude of added white noise to decompose the vibration signals into some IMFs.
2) To follow the steps 2-5 explained in section 5.1 in order to create normalized feature vectors and construct SVM classifier.

The bearing data set (acceleration signals) were collected using the test rig 1 (described in section 4.1). In addition to the healthy bearing, two defective bearings were utilized during the test, one with the very small artificial defect severity over one roller (150 microns in diameter) and another with the same fault level on the inner ring. A number of real vibration signals (various operating conditions and the two mentioned damage locations) is analyzed to verify effectiveness and robustness of the proposed method in discriminating and separating the faulty conditions.
The original acquired signals are divided into 20 segments including 10000 data points each, to extract required informative feature vectors. Thus, each signal includes 20 segments which create 20 feature vectors as inputs for the SVM.

Results and analysis

By performing the methodology, the feature vectors for each algorithm, damage location and signal direction are obtained. The normalized energy of IMFs introduced as an efficient feature vector in fault diagnosis of roller bearing, has been adopted using only first three elements of the feature vectors, i.e. IMF1 to 3.

In Table 5.8, it is shown that the smallest defect size (150 microns on a roller) is not recognized using EMD (in some working conditions) for signals collected through the accelerometer in Y direction (section 4.1). Although applying the EEMD (with the 100 trails) improves the success rates, it could not perfectly classify the samples in some operating conditions. It seems that amplitudes with 0.3, 0.5 and 0.6 lead to less misclassification (only in three operating conditions).

Now, by applying the PIEEMD algorithm, the accuracy of damage detection is investigated.

Table 5.8- The classification rates for different operating conditions (signals collected in Y direction)

Method	Noise	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
EMD	-	100	100	**75.0**	100	100	**81.3**	100	**81.3**	100	100	**93.8**	**62.5**
EEMD	0.2	100	100	**81.3**	100	**93.8**	**87.5**	100	**87.5**	100	100	100	100
EEMD	0.3	100	100	**93.8**	100	100	**87.5**	100	**93.8**	100	100	100	100
EEMD	0.4	100	100	**87.5**	100	**93.8**	**81.3**	100	100	100	100	100	100
EEMD	0.5	100	100	**87.5**	100	**87.5**	**87.5**	100	100	100	100	100	100
EEMD	0.6	100	100	**87.5**	100	**87.5**	**87.5**	100	100	100	100	100	100

The success rate of defect detection is shown in Table 5.9 for various preselected SNR values. Obviously, a considerable improved success rates are achieved for some SNR values, especially for SNR = 10 so that there exist only one working condition (speed = 200 Hz and load = 1.8 kN) whose bearing state is not perfectly identified. Increasing the SNR (to SNR = 20), increases the success rate for the mentioned operating condition which means that such a signal needs weaker noise to affect the extrema and consequently decrease the mode mixing. Whereas for some signals (such as speed = 100 Hz / load = 1.8 kN and speed = 300 Hz / load = 1.4 kN) the smaller value (SNR = 10) seems to be more appropriate. It means that those signals require some stronger noises. Exploring reliability of the obtained success rate, the Margin of each SVM classification is calculated and presented in Table 5.10. It is clear from the definition of Margin that higher Margin means more reliable hyperplane and classification. As it can be seen, some Margins are much smaller than others such as 0.6806 (SNR = 20, speed = 300 Hz and load = 1.4 kN). It means that the possibility of sorting out new misclassified samples (like the result shown in Table 5.9) is increased. On the other hand, it is reasonable to expect that higher Margins have more reliable results (a correct classification and defect detection for new unknown sample). The most important conditions are those achieving a perfect classification rate (100%), whereas the calculated Margin is not high enough such as 0.7749 (SNR = 20, speed = 200 Hz and load = 1.8 kN).

Table 5.9- The success rates of damage detection using the proposed algorithm (PIEEMD) in Y direction

SNR	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
5	100	100	**81.3**	100	**87.5**	**73.3**	100	100	100	100	100	100
10	100	100	100	100	100	**87.5**	100	100	100	100	100	100
15	100	100	**93.8**	100	100	**93.8**	100	**87.5**	100	100	100	100
20	100	100	**87.5**	100	100	100	100	**87.5**	100	100	100	100
25	100	100	**81.3**	100	100	**93.8**	100	**87.5**	100	100	100	**93.8**
30	100	100	**81.3**	100	100	**93.8**	100	**87.5**	100	100	100	**93.8**

Table 5.10- The Margin calculated using the proposed algorithm (PIEEMD) in Y direction

SNR	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
5	1.1553	1.0479	**0.7044**	1.2654	**0.7560**	**0.7179**	1.2676	1.0039	0.9007	0.8362	1.0402	0.9647
10	1.1627	1.0881	0.8863	1.2660	0.8095	**0.6854**	1.1806	0.8123	0.8628	0.9342	0.8768	0.9274
15	1.2295	1.1278	**0.7577**	1.2779	0.8885	**0.7732**	1.1139	**0.7493**	0.8810	0.9128	0.8454	0.8652
20	1.2519	1.1192	**0.7303**	1.2750	0.9342	0.7749	1.0344	**0.6806**	0.8959	0.8668	0.8547	0.7864
25	1.2066	1.1015	**0.6819**	1.3166	0.9003	**0.7793**	1.0172	**0.7014**	0.9039	0.8631	0.8189	**0.7514**
30	1.1962	1.0799	**0.7069**	1.2634	0.8601	**0.7705**	0.9985	**0.7425**	0.9084	0.8292	0.7611	**0.7108**

Table 5.11- The Margin calculated using EEMD with different amplitudes (Y direction)

Method	Noise	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
EMD	-	1.1684	1.0417	0.6961	1.2276	0.8157	0.7156	0.9656	0.7523	0.8620	0.8438	0.7013	0.7109
EEMD	0.3	1.2248	1.1168	0.8136	1.3011	0.8000	0.6922	1.2131	0.8100	0.9289	1.0498	0.8895	0.9143
EEMD	0.5	1.2162	1.1324	0.8112	1.3009	0.7000	0.6372	1.2868	1.0633	0.9100	0.8995	0.8779	0.9069
EEMD	0.6	1.2388	1.1706	0.7633	1.3019	0.7693	0.7099	1.2880	1.0276	0.9312	0.9056	0.8980	0.9543
SNR	10	1.1627	1.0881	0.8863	1.2660	0.8095	0.6854	1.1806	0.8123	0.8628	0.9342	0.8768	0.9274
SNR	20	1.2519	1.1192	0.7303	1.2750	0.9342	0.7749	1.0344	0.6806	0.8959	0.8668	0.8547	0.7864

Table 5.12- The reliability test of the constructed SVM with 20 new samples (Y direction)

Method	Noise	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
EMD	-	100	100	**75**	100	100	**80**	100	**85**	100	100	**85**	**65**
EEMD	0.3	100	100	**90**	100	100	**80**	100	**90**	100	100	100	100
EEMD	0.5	100	100	**90**	100	**85**	**75**	100	100	100	100	100	100
EEMD	0.6	100	100	**85**	100	**85**	**80**	100	100	100	100	100	100
PIEEMD (SNR)	10	100	100	100	100	100	**80**	100	100	100	100	100	100
PIEEMD (SNR)	20	100	100	**80**	100	100	**90**	100	**80**	100	100	100	100

Since there exist the possibility of misclassification for the new samples based on the constructed SVM. It is worth to mention that it does not exist a determined reliable value for Margin. The Margin of those amplitude calculation algorithms achieve better results are shown in Table 5.11. In order to test the reliability of the constructed SVM, 20 new samples (10 healthy and 10 damaged samples) for each operating condition are classified with the previously constructed SVM (Table 5.9). The results are proposed in Table 5.12. Obviously, as it was expected, the new samples are not classified perfectly for previously mentioned low Margin (SNR = 20, speed = 200 Hz and load = 1.8 kN). Although the previous success rate was 100% and it seemed to be a reliable constructed SVM, its low Margin (in comparison with those showing a perfect damage detection) indicates that it may not be a confident SVM (Fig. 5.11). However, for all other conditions, which have higher Margin, the states of the new samples are correctly identified. As it can be seen in Fig. 5.12, all faulty and healthy samples are completely separable.

For the EMD and EEMD (with the predefined constant amplitudes or the presented method (SNR = 10), the previously constructed SVMs can successfully detect the fault, as it was expected, because of their high Margins. The results shown in Table 5.12 confirm such an expectation.

In order to investigate in more details, the collected signals of another defective bearing (small defect on the inner ring) in two directions (Y and Z) are analyzed. The results of classification are shown in Table 5.13. As all the SVMs constructed have high Margins, they are reliable and leads to perfect success rates for both Y and Z directions, except for one condition (speed = 200 Hz and load = 1.8 kN) in Z axis which achieved 81.3% success rate.

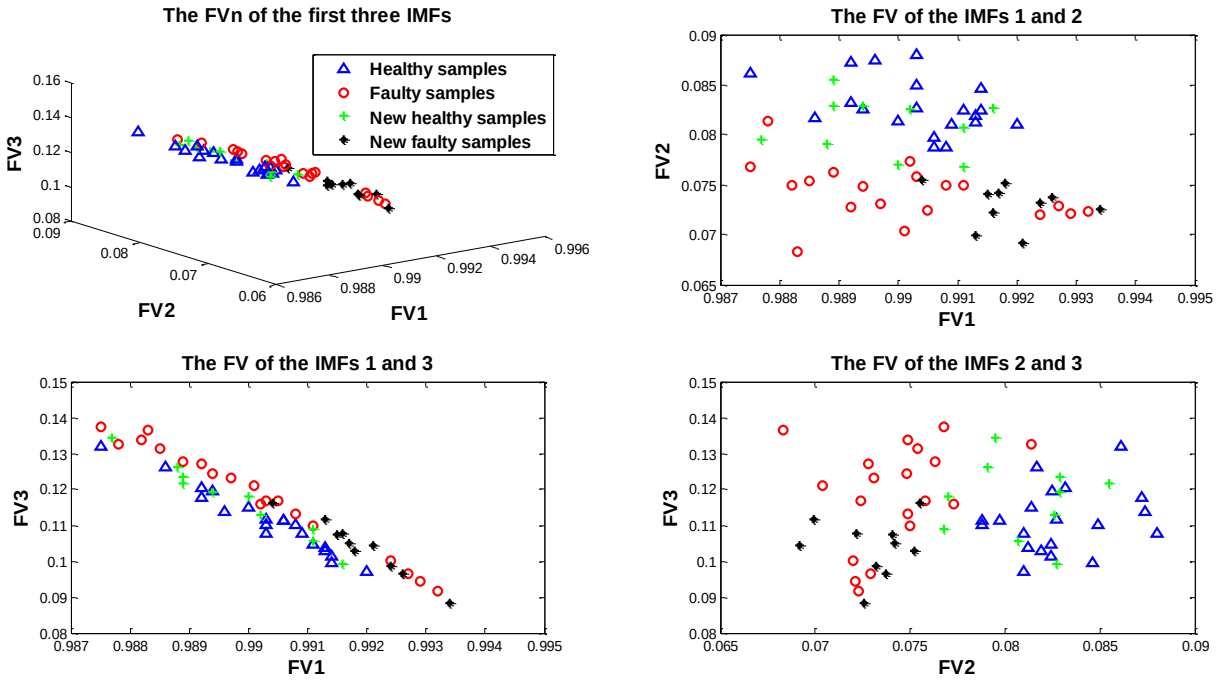


Figure 5.11- The normalized energy of the three first IMFs ($\mathbf{FV}_n$) with new samples, SNR = 20 (speed = 200 Hz and load = 1.8 kN)

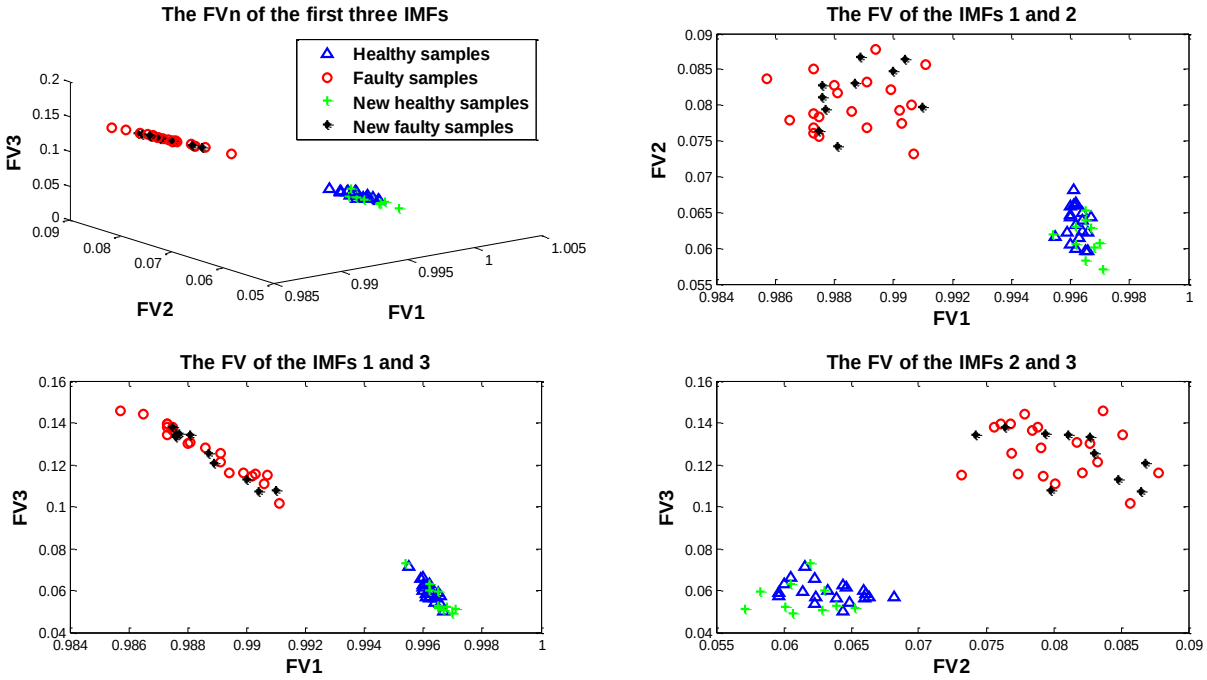


Figure 5.12- The normalized energy of the three first IMFs ($\mathbf{FV}_n$) with new samples, SNR = 10 (speed = 200 Hz and load = 1.0 kN)

Finally, the appropriate ensemble trial number for damage detection is investigated. The results of fault diagnosis for different trials are presented in Table 5.14. In the most of the operating conditions, the 30 ensemble trial numbers is enough to achieve a perfect labelling rate, which is far lower than the numbers used in previous studies. It leads substantially to less time consuming and decreasing the computational cost. The higher Margin is not obtained by increasing the number, even in some cases the 30 trials achieves a little higher Margin. For two conditions (Speed = 200 Hz, load = 1.4 kN and speed = 300 Hz, load = 1.4 kN), the 100% success in detection is achieved by applying the 100 trails. However, there exists only one operating condition in which the 100 trials does not achieve perfect success rate (Speed = 200 Hz and load = 1.8 kN). As it can be seen in Table 5.15, implementing the proposed method with 50 trials improves the success rate and the Margin. Increasing the trials to 1000, a little higher classification accuracy is achieved (from 87.5% to 93.8%), whereas the Margin does not change considerably after the 50 trial numbers. Although with the 70 trials the classification result improves from 81.3% to 87.5%, the original EEMD does not obtain higher rates by applying more trial numbers (even up to 1000).

On the contrary, denoising the signals can improve the results to obtain perfect classification results (section 5.2). Using only 30 trials, the success rate is 93.8%, whereas the same rate was achieved with 1000 trials for the noisy signals (Table 5.16). The 50 trials seems to be sufficient to achieve 100% labelling classification rate. Furthermore, increasing the trail could not improve considerably the Margins.

Table 5.13- The Margin calculated (using PIEEMD with SNR = 10), for the defective inner ring

Direction	100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
Y	1.2186	0.8674	1.4268	1.3744	1.2852	1.0538	1.3522	1.1211	0.8542	1.2552	1.2583	0.8326
Z	1.2603	0.8788	1.3713	1.3391	0.9743	**0.7350**	1.2272	1.0937	1.1128	1.1918	1.1989	1.3516

Table 5.14- The results of using different trial numbers (NR) using PIEEMD (SNR 10) for Y direction

NR		100 Hz 1.0 kN	100 Hz 1.4 kN	100 Hz 1.8 kN	200 Hz 1.0 kN	200 Hz 1.4 kN	200 Hz 1.8 kN	300 Hz 1.0 kN	300 Hz 1.4 kN	300 Hz 1.8 kN	400 Hz 1.0 kN	400 Hz 1.4 kN	400 Hz 1.8 kN
100	Margin	1.1627	1.0881	0.8863	1.2660	0.8095	0.6854	1.1806	0.8123	0.8628	0.9342	0.8768	0.9274
	Success rate	100	100	100	100	100	87.5	100	100	100	100	100	100
90	Margin	1.1584	1.0700	0.8506	1.2383	0.7776	0.7356	1.1625	0.8339	0.8704	0.9009	0.8886	0.9049
	Success rate	100	100	100	100	93.8	93.8	100	93.8	100	100	100	100
70	Margin	1.1714	1.0856	0.8575	1.2234	0.7344	0.7043	1.1347	0.7825	0.8819	0.8238	0.8668	0.9005
	Success rate	100	100	100	100	93.8	87.5	100	93.8	100	93.8	100	100
50	Margin	1.1487	1.0633	0.8858	1.2563	0.7450	0.7244	1.1732	0.8410	0.8700	0.8748	0.8685	0.9103
	Success rate	100	100	100	100	93.8	87.5	100	93.8	100	100	100	100
30	Margin	1.1586	1.1026	0.8378	1.2413	0.8442	0.6353	1.1698	0.8271	0.9037	0.8861	0.9065	0.9346
	Success rate	100	100	100	100	93.8	81.3	100	87.5	100	100	100	100

Table5. 15- The results of using different trial numbers (NR) for the condition: speed = 200 Hz and load = 1.8 kN

Method		NR						
		30	50	70	100	500	1000	2000
PIEEMD (SNR 10)	Success rate	81.3	87.5	87.5	87.5	87.5	93.8	93.8
	Margin	0.6353	0.7244	0.7043	0.6854	0.7254	0.7151	0.7267
EEMD (0.3)	Success rate	81.3	81.3	87.5	87.5	87.5	87.5	87.5
	Margin	0.6912	0.6998	0.7032	0.6922	0.6975	0.7086	0.7116

Table 5.16- The results of using different trial numbers (NR) after denoising, for the condition: speed = 200 Hz and load = 1.8 kN

SNR 10	NR				
	30	50	70	100	500
Margin	0.7175	0.8119	0.7573	0.7643	0.7913
Success rate	93.8	100	100	100	100

Conclusions

It is shown that the proposed method (performance improved EEMD (PIEEMD)) achieves higher damage detection success rate and creates larger Margin than the original algorithm. Exploring the reliability, the Margin of each SVM classification is calculated and it is confirmed that for those conditions whose Margin is relatively high, the results are more reliable. For the defective inner ring, the acceleration signals of two radial directions are investigated to achieve more confident results and it is validated that the PIEEMD algorithm looks reliable and can be favorably applied instead of the previous pre-determined approach.

It is shown that substantially low trial numbers are required to achieve perfect labelling of samples, which leads to considerably less computational cost. In the most operating conditions only the 30 trial numbers are sufficient to achieve successful damage detection. Using the 100 trial numbers leads to 100% success rate for all working conditions, except for one condition (Speed = 200 Hz, load = 1.4 kN), which increasing the number improves its result, however, the perfect classification can be achieved only after denoising.

5.4 Anomaly detection using novel feature extraction

As shown in section 5.2 and 5.3, there exist still some cases for which the state of the bearing cannot perfectly identified.

In this study a new feature extraction method is proposed for anomaly detection of roller bearings. As Teager-Kaiser energy operator (TKEO) technique, discussed in section 2.2.3, can track the energy of signals, at any instant, it is applied to create feature vectors. The three dimensional feature vectors of healthy bearing signals are used to construct the one-class SVM. Since in some real applications, the faulty bearings data are not available.

Methodology

The goal of this study is to evaluate the performance of the proposed feature extraction algorithm in fault detection of a roller bearing.
The fault diagnosis method for the traditional EEMD technique is as explained in section 5.2. Whereas the proposed feature extraction is implemented as the following steps:

1. To decompose the signal using the PIEEMD with SNR = 10 dB (proposed in section 5.3).
2. To apply the TKEO to the first m IMFs of each signal.
3. To calculate the sum of each TKEO.

$$TKEO_i = \sum_{i=1}^{m} \psi(IMF_i) \tag{5.17}$$

4. To create a feature vector with the sum of the calculated TKEO:

$$TKE = [TKEO_1, TKEO_2, \dots, TKEO_m] \tag{5.18}$$

5. To normalize the feature:

$$TKE_n = [TKEO_1/TKEO_{tot}, TKEO_2/TKEO_{tot}, \dots, TKEO_m/TKEO_{tot}] \tag{5.19}$$

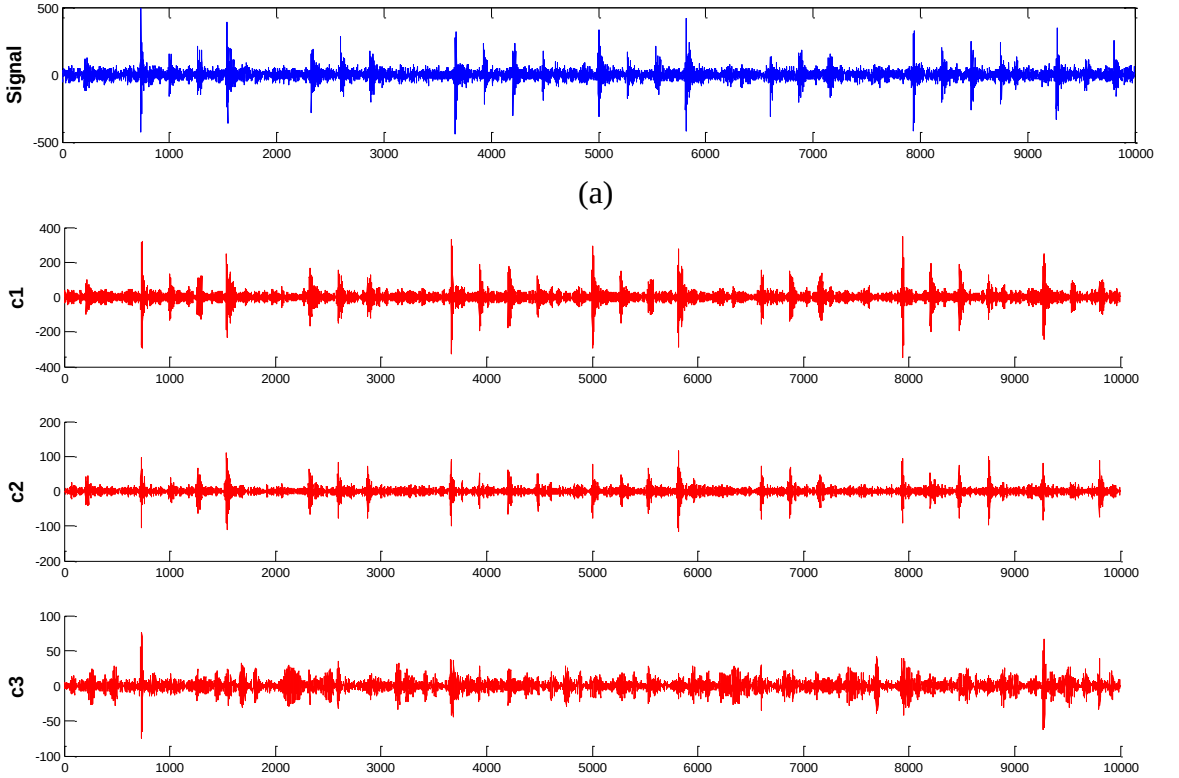
where $TKEO_{tot} = (\sum_{i=1}^{m} TKEO_i)$.

Finally, the training procedure of one-class SVM is carried out by utilizing the normalized feature vectors so far obtained. The 80% of healthy samples are used for training and the rest (remaining healthy samples and all faulty data) are taken as the test samples.

The bearing data set (acceleration signals) were collected using the test rig 1 (section 4.1). In addition to the healthy bearing, One defective bearing was utilized during the test with the very small artificial defect severity over one roller was 150 microns in diameter. A number of real vibration signals (various operating conditions) is analysed to verify effectiveness and robustness of the proposed method in discriminating and separating the faulty conditions. The original acquired signals are divided into 20 segments including 10000 data points each, to extract required informative feature vectors. Thus, each signal includes 20 segments which create 20 feature vectors as inputs for the SVM.

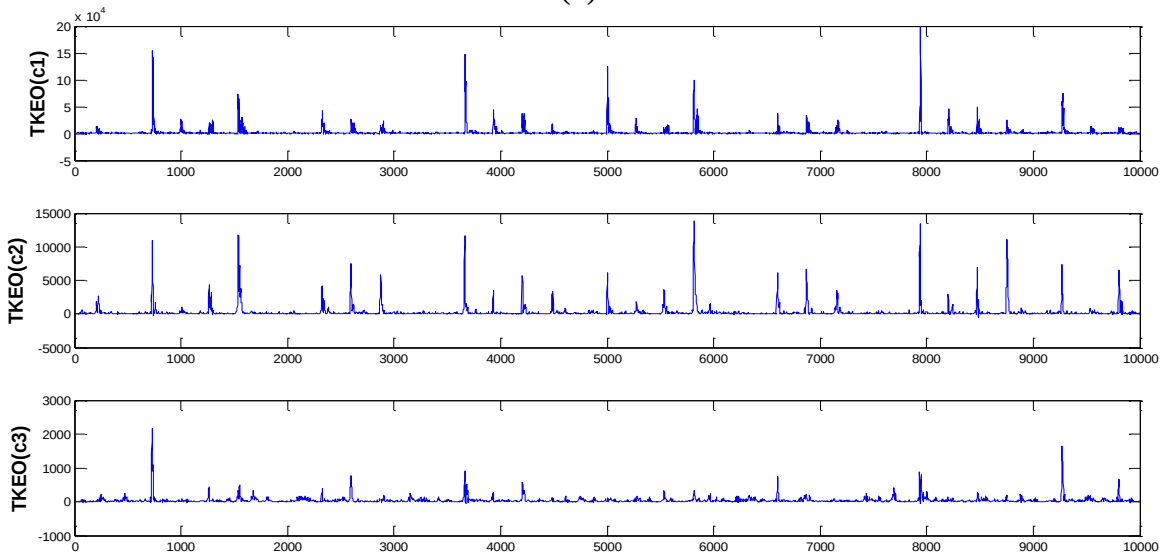
Results and analysis

An acquired acceleration signal, its three first IMFs (obtained using EEMD) and the TKEO of those IMFs are shown in Fig. 5.13. By implementing the methodology to the signals collected in tangential (Z) direction, the normalized energy (FV_n) of IMFs for the EEMD method and the normalized feature (TKE_n) for the proposed method are obtained. The 0.3 of standard deviation of each original signal is used as the appropriate amplitude of added noise in the original EEMD.



(a)

(b)



(c)

Figure 5.13- a) A collected acceleration signal, b) first three IMFs and c) TKEO of those IMFs.

As it can be seen in Fig. 5.14, there exist a confusion among healthy and faulty samples obtained by the EEMD method. In view of this, the novel feature proposed along this section is applied to check whether it can improve the performances of detection. As it is evidenced in Fig. 5.15, the healthy and faulty samples are perfectly separable. Thus, it is expected to achieve higher success rate in labelling of new samples.

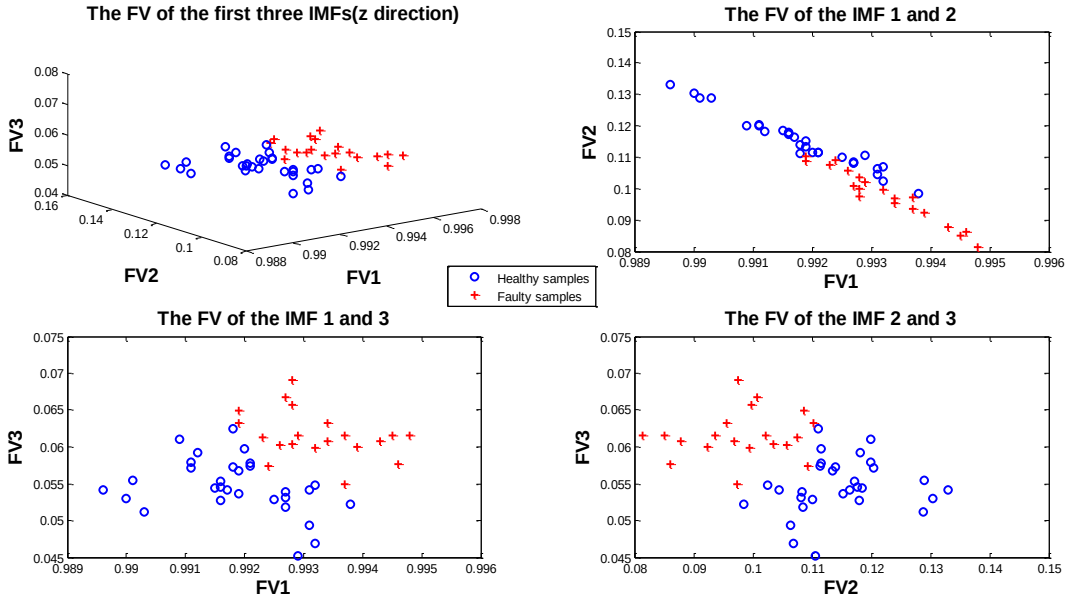


Figure 5.14- The feature vector of three first IMFs (FV_n) using EEMD (Speed = 300 Hz and load = 1.8 kN)

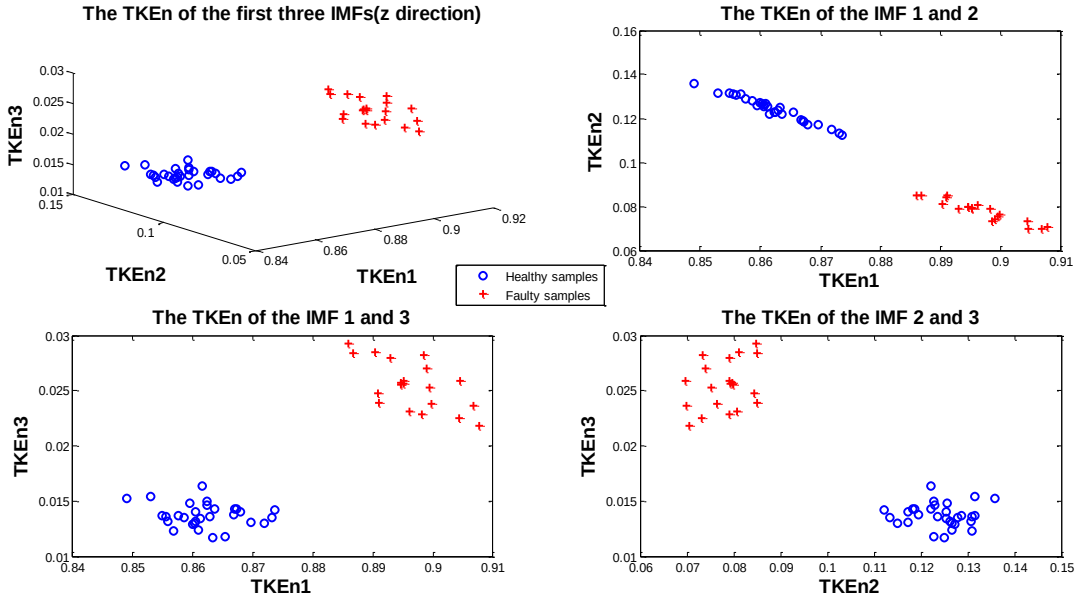


Figure 5.15- The novel feature vector of three first IMFs (TKE_n) (Speed = 300 Hz and load = 1.8 kN)

In Tables 5.17 and 5.18, the results of classification (For shaft speed = 200 and 300 Hz) using one-class SVM are shown. The results are highly dependent on the classification parameters. The optimal values of the classification parameters (γ and ν) are presented for each methods. The success rates obtained using the proposed feature extraction, are higher so that in some cases there exist considerable differences. For example, in the condition 300 Hz speed and 1.8 kN load, the proposed technique improves the test success rate up to 23.1%.

In addition to the classification rate index, the complexity of the constructed hyperplane and the calculated Margin are used to compare the results from traditional EEMD and the proposed method (Tables 5.19 and 5.20). The bearing condition can be perfectly recognized (using EEMD) for a single working condition (Speed = 200 Hz and load = 1.4 kN). In this condition, the fraction of Support Vectors (SVs) is 8/24, whereas applying the new method the complexity of the hyperplane decreases because it is defined by a lower SVs fraction (5/24). Furthermore, the Margin created by the EEMD is 0.999305, while using the PIEEMD the Margin is improved to 1.146190.

Table 5.17- The classification results for both methods (Shaft speed = 200 Hz)

Method	Load											
	1.0 kN				1.4 kN				1.8 kN			
	γ	ν	training	test	γ	ν	training	test	γ	ν	training	Test
EEMD	0.3	0.1	100	96.2	0.1	0.3	100	100	0.3	0.05	100	92.3
PIEEMD+TKEO	0.1	0.3	100	100	0.1	0.3	100	100	0.1	0.3	100	100

Table 5.18- The classification results for both methods (Shaft speed = 300 Hz)

Method	Load											
	1.0 kN				1.4 kN				1.8 kN			
	γ	ν	training	test	γ	ν	training	test	γ	ν	training	Test
EEMD	0.1	0.3	95.8	96.2	0.1	0.1	95.8	92.3	0.05	0.3	100	73.1
PIEEMD+TKEO	0. 05	0.35	100	100	0.05	0.2	100	92.3	0.1	0.2	100	96.2

Table 5.19- The fraction of SVs and calculated Margin (Shaft speed = 200 Hz)

Method	Load					
	1.0 kN		1.4 kN		1.8 kN	
	Fraction of SVs	Margin	Fraction of SVs	Margin	Fraction of SVs	Margin
EEMD	3/24	0.999722	8/24	0.999954	2/24	0.999305
PIEEMD+TKEO	8/24	1.13989	8/24	1.16538	8/24	1.26256

Table 5.20- The fraction of SVs and calculated Margin (Shaft speed = 300 Hz)

Method	Load					
	1.0 kN		1.4 kN		1.8 kN	
	Fraction of SVs	Margin	Fraction of SVs	Margin	Fraction of SVs	Margin
EEMD	8/24	1.000010	3/24	0.999994	8/24	1.000030
PIEEMD+TKEO	8/24	1.098540	5/24	1.094330	5/24	1.146190

It means that the proposed feature extraction generates a less complex and more reliable hyperplane.

In all operating conditions, by adopting the proposed method to construct the hyperplane, higher Margins are obtained, which indicates more reliable classification. It achieves the perfect success rates in the most cases, except for two operating conditions (Speed = 300 Hz, load = 1.4 and 1.8 kN). Even in these conditions, the success rates are higher than the EEMD. In the load = 1.8 kN condition, there exist only one misclassified sample, which is a healthy sample labelled as a faulty bearing (faulty alarm). In fault diagnosis, it is more important not to classify a faulty sample as a healthy one than having a faulty alarm. When the ν parameter approaches zero, the problem then resembles the corresponding hard margin algorithm, since the penalization of errors becomes infinite (section 3.3). As it can be seen in Tables 5.17 to 5.20, in some cases, the constructed hyperplane based on EEMD, seems to be hard-margin because of very low ν and very small number of SVs. For example, the condition corresponding to the speed of 200 Hz and the applied load of 1.8 kN, the parameter value is 0.05 and the achieved number of SVs is only 2. It indicates a hard-margin condition that only a few outlier can determine the boundary and makes the classifier significantly sensitive to noise in the data. By increasing the ν parameter to create a soft-margin model, the training accuracy will be reduced considerably.

In contrast, all the constructed models based on the proposed feature extraction method are soft-margin SVM and more reliable.

In order to detect the larger defect size (450 microns), the proposed feature extraction method is applied and the perfect success rates of classification are achieved for all operating conditions. As it can be seen in Fig. 5.16, the healthy and faulty samples are perfectly separable, even for the condition where the states of the bearing were not detected perfectly for the smaller defect size (Speed = 300 Hz and load = 1.4 kN).

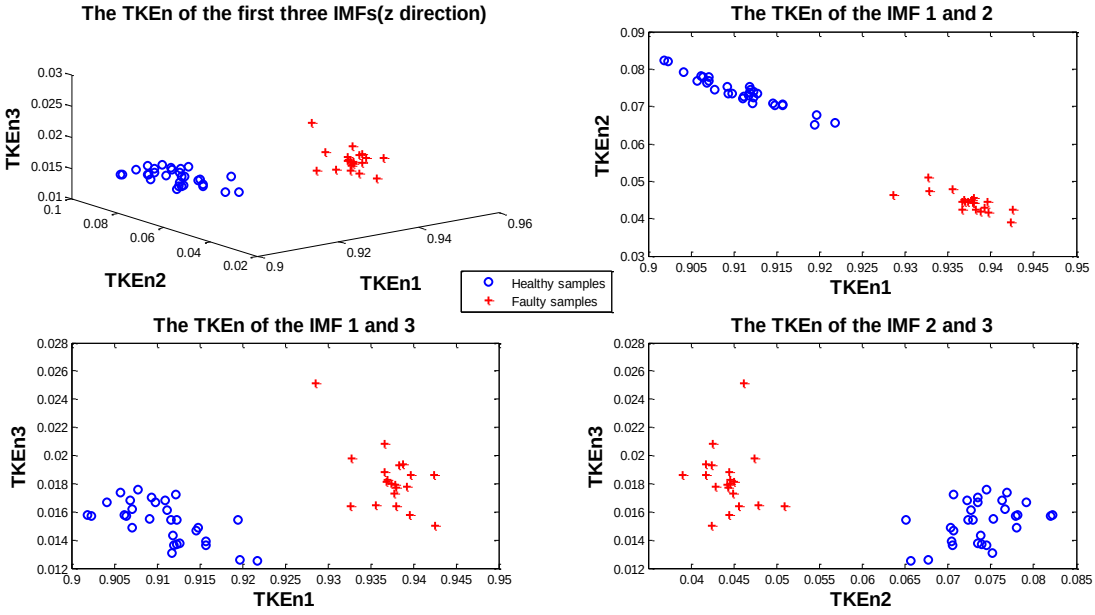


Figure 5.16- The novel feature vectors of three first IMFs ($\mathbf{TKE}_n$) for the 450 microns defect size

Conclusions

Applying the EEMD does not lead to a perfect anomaly detection in the case of small size defect (150 microns). However, it is shown that the proposed feature extraction method is a powerful method for detecting even the smallest damage level (150 microns) so that it can classify the samples perfectly in many operating conditions. It create less complex (less fraction of SVs) and more reliable (higher Margin) hyperplane than the EEMD method. For the larger defect size (450 microns), utilizing the proposed technique, the healthy and faulty samples are completely separable and the success rates of labelling the new samples are exact in all operating condition.

5.5 Defect-sensitive feature for time-varying operational conditions

The diagnostic methods shown in this thesis seem to be powerful for only stationary conditions (constant speed and load). However, bearings often work at time-varying conditions such as for wind turbine supporting bearings, mining excavator bearing, vehicles, robots and all processes with run-up and run-down transients. Extracting feature for those nonstationary operating conditions, especially for early defects, requires an appropriate defect-sensitive method which, at the same time, is insensitive to operational and environmental condition variations. The order tracking method has been widely used when there exist speed variations. However, there still exist two drawbacks: limitation of speed variation and measurement cost. It can effectively detect faults when the speed variation is limited. Moreover, an extra device such as an encoder or a tachometer is required to provide a phase reference signal, which increases instrumentation cost and produces installation and adjustment problems. Many studies have been carried out to improve it [64-69] and to develop it so that it can be applied in tacho-less form [267,268].

In this study, a new method is proposed and investigated to detect the state of roller bearings operating in time-varying condition based on cointegration. The method decomposes the acquired signal into intrinsic mode functions (IMFs), then cointegration (section 2.2.4) is subsequently applied to the obtained intrinsic mode functions to extract stationary residuals, which are insensitive to operating conditions. It is looking for linear stationary combinations within non-stationary time series.

The bearing data set (acceleration signals) were collected using the test rig 2 (section 4.2) assembled at the Department of Mechanical and Aerospace Engineering of Strathclyde University. The small artificial defect was created on the inner ring, which is illustrated in section 4.2. A run-up time-varying operating condition was considered to acquire the signals coming from the healthy and faulty bearings. The measurement was manually started from 150 rpm to 1500 rpm and each signal includes 30000 elements. The data consists of 20 measured signals for each of the healthy and faulty bearings. One acquired acceleration signal for normal and one for damaged bearings are shown in Fig.5.17. The signal is divided into segments and each segment is broken down into some elementary modes the intrinsic Mode Functions (IMFs) using the performance improved EEMD (section 5.4)Then, three dimensional feature vectors are created by applying the Teager-Kaiser Energy Operator (TKEO) to the cointegrated residuals of the first three IMFs. The feature vectors obtained from healthy bearing signals are further utilized as input to construct a separating hyperplane for a one-class Support Vector Machine (SVM). The SVM can be trained to categorize signals coming from healthy and faulty bearings.

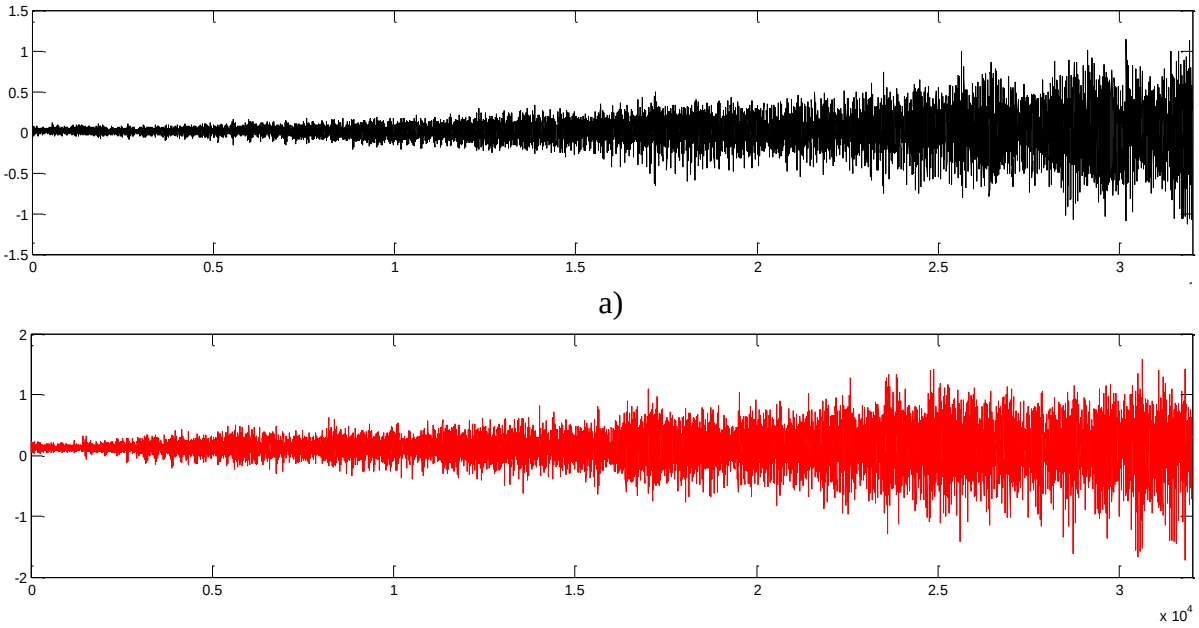


Figure 5.17- Two acquired signals during run-up condition a) from the healthy bearing b) from the damaged bearing.

Methodology

The method proposed is given as the following:

1. To divide each collected signal into five equal segments so that each segment includes 6000 elements.
2. To apply the PIEEMD to each segment, in order to decompose them into some IMFs.
3. To Select first three IMFs of each segment and call them IMF1, IMF2 and IMF3, respectively.
4. To test the existence of cointegrated residuals for the five first IMF1s from the segments of the whole signal.
5. To apply the step 4 for the five IMF2 and five IMF3 from the five signal segments.
6. To apply the TKEO to the obtained cointegration vectors obtained from the five IMF1, IMF2 and IMF3 and calculate the sum of the obtained TKEOs as shown below.

$$TKE_i = \sum \psi(cointegration\ vector(IMF_i)) \tag{5.20}$$

where ψ denotes the Teager-Kaiser operator.

7. To create a three dimensional feature vector.

$$FV = [TKE_1 \quad TKE_2 \quad TKE_3] \tag{5.21}$$

where TKE_1, TKE_2, and TKE_3 are calculated for IMF_1, IMF_2 and IMF_3 using Eq. 5.20, respectively.

8. To normalize the feature vectors by dividing to the norm.

$$FVn = [TKE_1/norm(TKE_1) \quad TKE_2/norm(TKE_2) \quad TKE_3/norm(TKE_3)] \tag{5.22}$$

which can be written as:

$$FVn = [FV_1 \quad FV_2 \quad FV_3] \tag{5.23}$$

9. To utilize the healthy feature vectors to build the classifying hyperplane by one-class SVM. The 80% of healthy samples are used for training and the rest (remaining healthy samples and all faulty data) are taken as the test samples. Once the training procedure is successfully performed, the parameters are hold to test samples to identify the different work conditions and fault patterns.

Results and analysis

First, the novel feature extraction proposed in section 5.4 is applied to the signals collected in time-varying operating condition (run-up). The result of applying this method is shown in Fig. 5.18. Although the feature extraction method is a powerful method for small fault level detection, it cannot identify whether the bearing is healthy or faulty when it is working in non-stationary condition. The healthy and faulty vectors are not separable but are significantly confused. Now, it is investigated if the method proposed in this study (explained in the section of Methodology) can recognize the state of the bearings. First, each signal is divided into five segments so that each segment includes 6000 data points. Then, each segment is decomposed by the PIEEMD algorithm (as it is displayed for a healthy and faulty signals in Figs. 5.19 and 5.20). The cointegration procedure is carried out for each five common number of IMFs to achieve one or more cointegrating vectors.

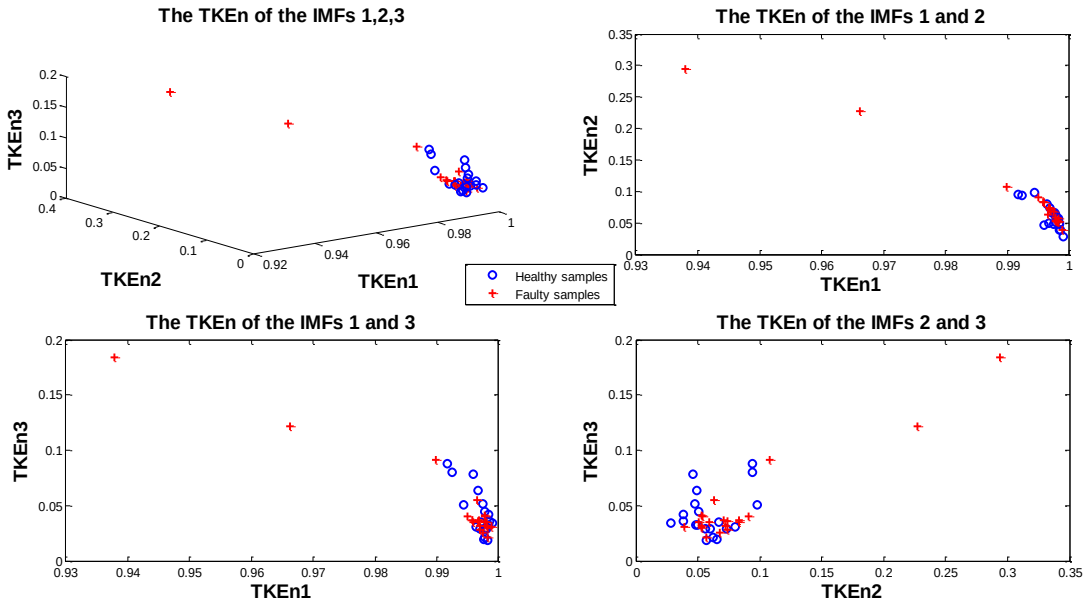


Figure 5.18- The normalized TKE of three first IMFs ($TKEn$) using PIEEMD.

For example, the five IMF1 (first IMFs of five segments shown in Fig. 5.21) are used to obtain the possible cointegrating vectors. The cointegration residuals for these IMFs are displayed in Fig. 5.22 and the results are given in Table 5.21. As it can be seen, the null ranks are displayed in the first column. The Cointegration ranks labeled r0, r1, etc. The values of h equal to 1 indicate rejection of the null hypothesis, which means there is a cointegration vector. It is full ranked since all the null hypothesis are rejected. The right-tail probabilities of the test statistics are considerably less than the significance level, which is 0.05. The test statistics value decreases from r0 to r4. The critical values for right-tail probabilities are presented determined as explained in section 2.2.4. The eigenvector corresponding to the largest eigenvalue is the most stationary cointegration vector. Thus, the most stationary cointegration vector is the first one (r0) and the stationarity degree decreases from r0 to r4.

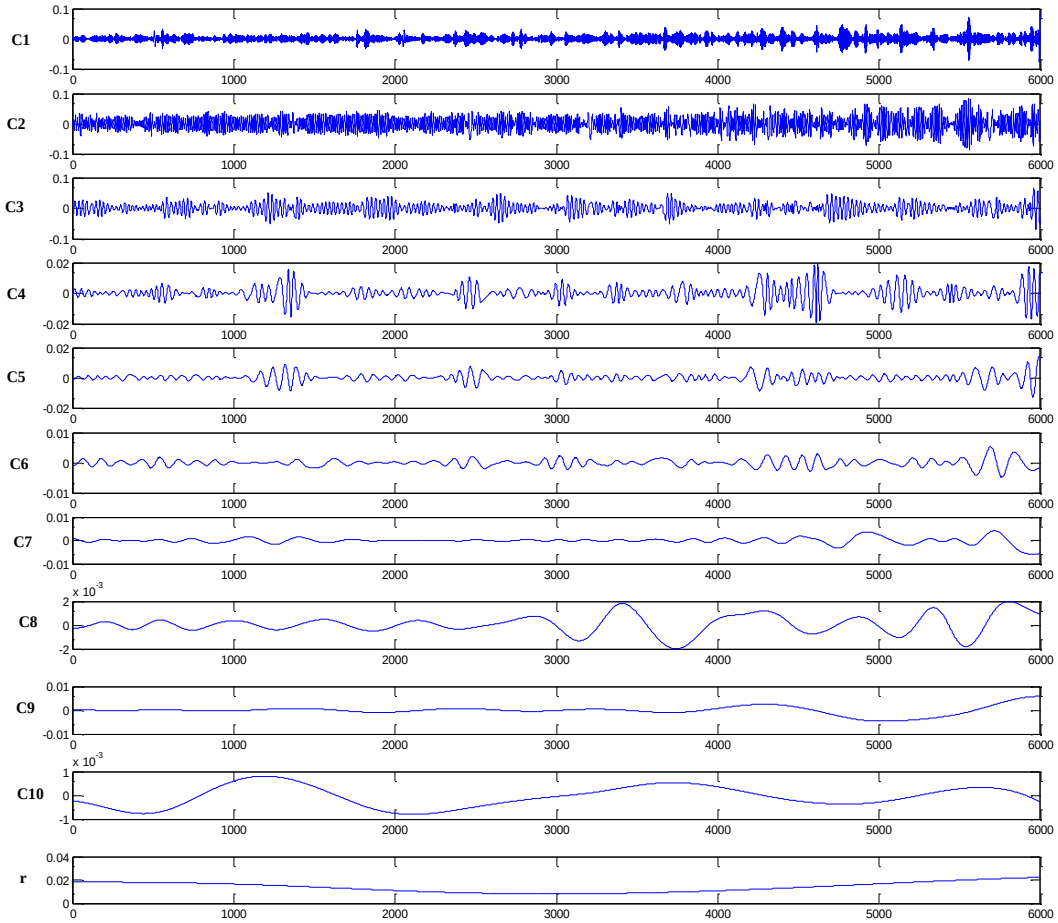


Figure 5.19- the IMFs of the healthy signal (Fig. 5.17a) obtained by performance improved EEMD (PIEEMD).

Finally, the TKEO is applied to each cointegrating vector (Eq. 5.20) and the results are shown in Fig. 5.23. The three dimensional feature is created using Eq. 5.22. After normalizing the feature vectors are used for the classification. Thus, for each signal divided into five segments, we achieve one three dimensional feature vector.

The healthy and faulty feature vectors are shown in Figs. 5.24 to 5.27 for each cointegrating vectors. The normal and damaged samples of the first and the second cointegrating vectors shown in Figs. 5.24 and 5.25 are obviously separable. The feature vectors extracted from healthy signals used as training input of the one-class SVM. Then, the hyperplane constructed is utilized for labelling the new samples, test samples includes faulty samples and 20% of healthy samples.

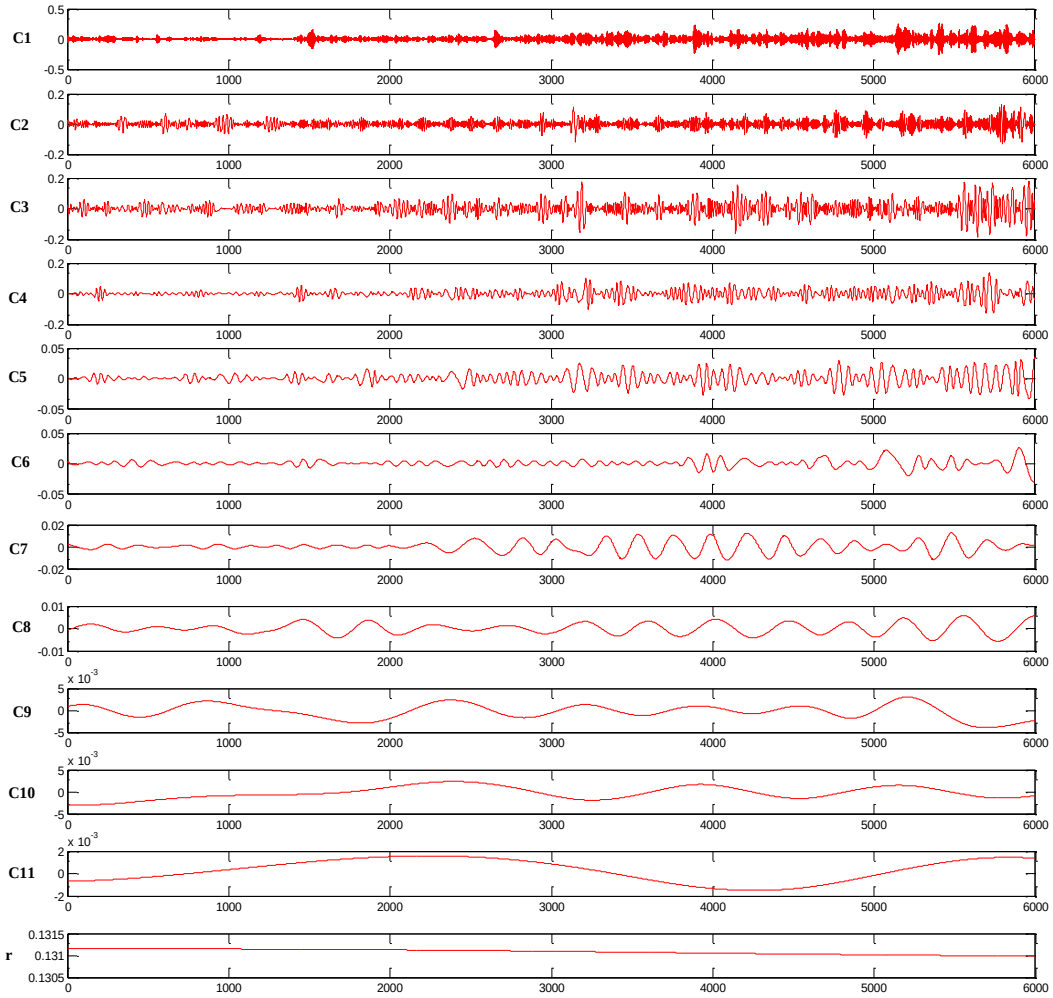


Figure 5.20- the IMFs of the faulty signal (Fig. 5.17b) obtained by performance improved EEMD (PIEEMD).

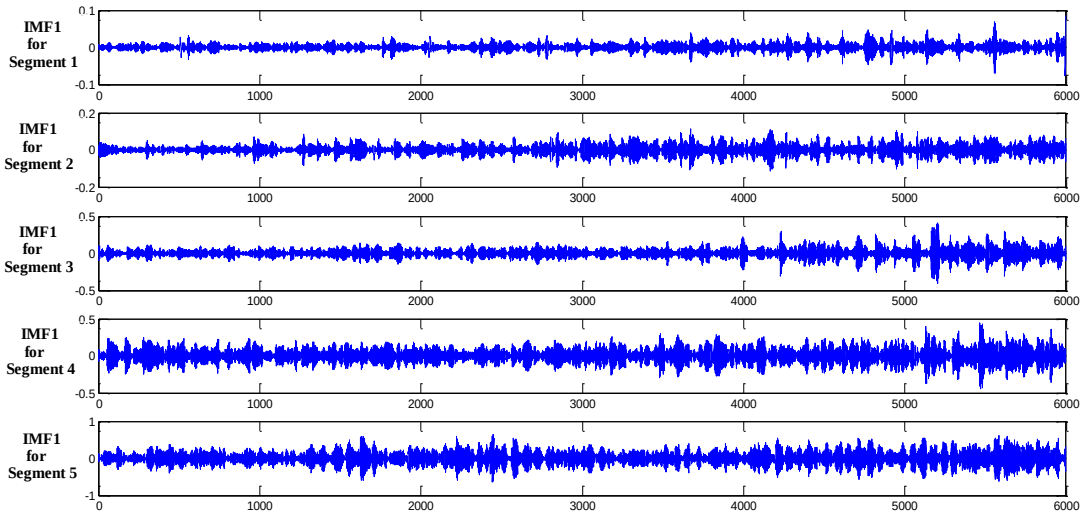


Figure 5.21- The first IMFs (IMF1) of the five segments of the signal shown in Fig. 5.17a.

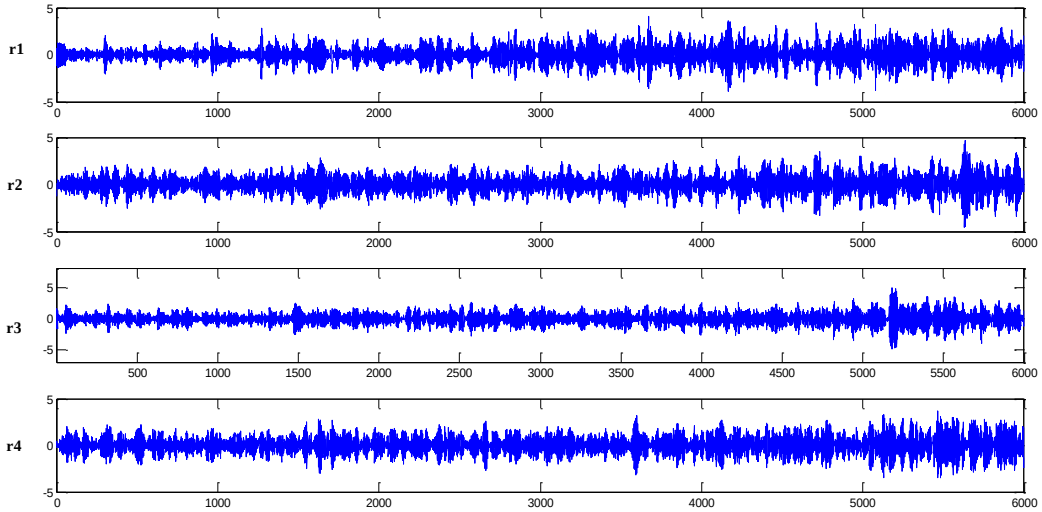


Figure 5.22- The cointegrating vectors obtained from IMF1 of the five segments shown in Fig. 5.21.

Table 5.21- The results of cointegration for IMF1 of five segments of the healthy signal shown in Fig. 5.21

Null ranks r	Null hypothesis h	Test statistics	Critical values	Right-tail probabilities	Eigen value
0	1	24466.1675	69.8187	0.0010	0.5753
1	1	19328.1531	47.8564	0.0010	0.5604
2	1	14397.1423	29.7976	0.0010	0.5589
3	1	9486.5483	15.4948	0.0010	0.5565
4	1	4608.9054	3.8415	0.0010	0.5362

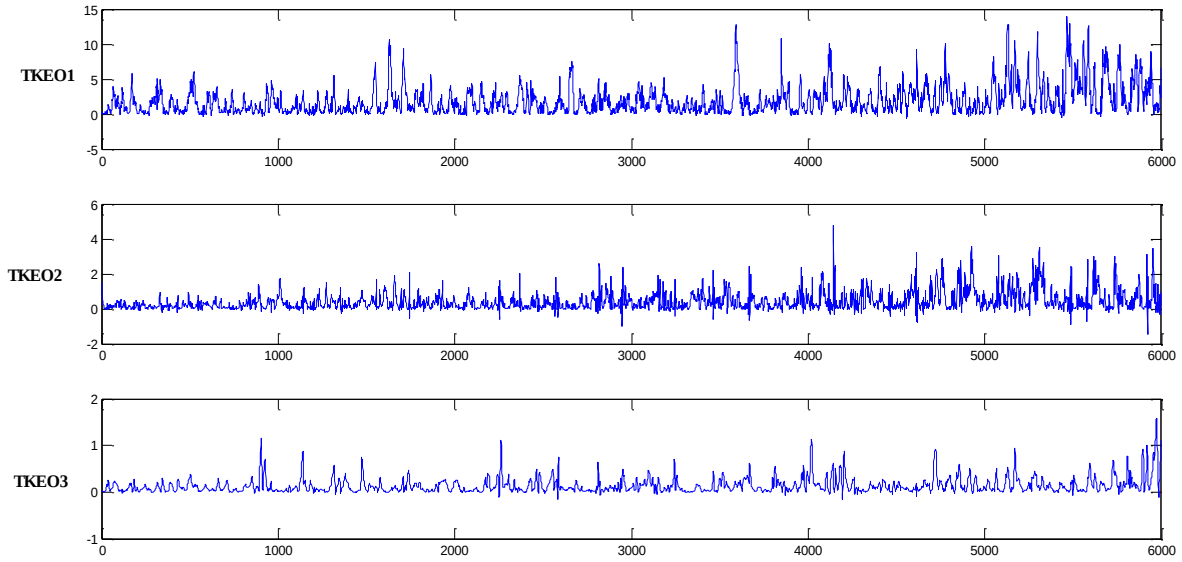


Figure 5.23- The TKEO applied to the cointegration vectors obtained from the five IMF1, IMF2 and IMF3, respectively.

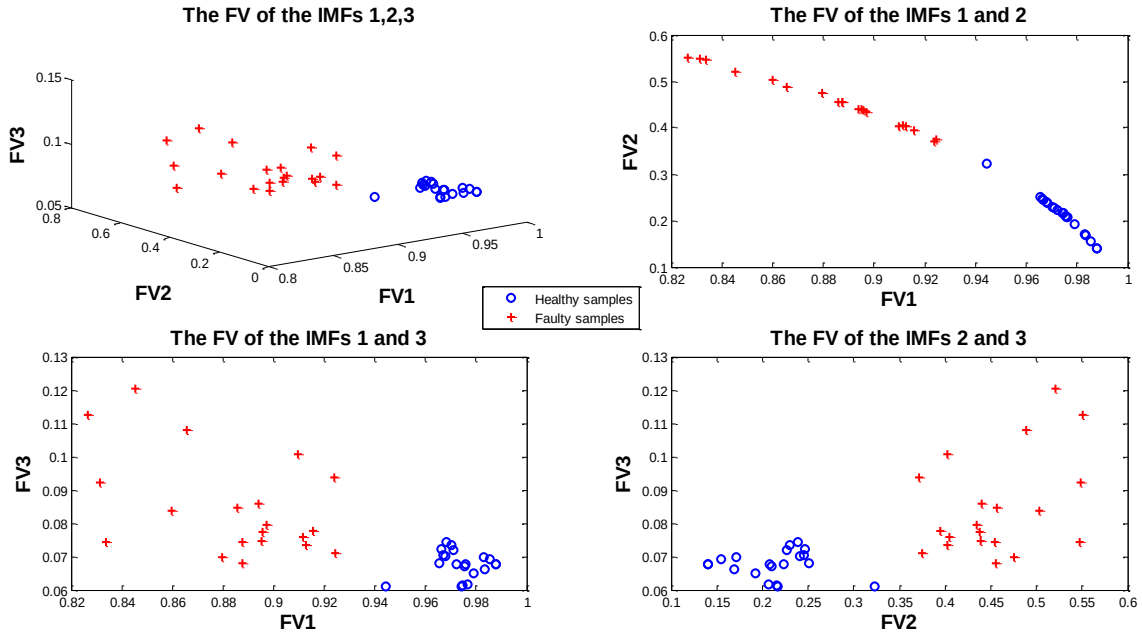


Figure 5.24- The three dimensional feature vectors obtained from first cointegrating vectors.

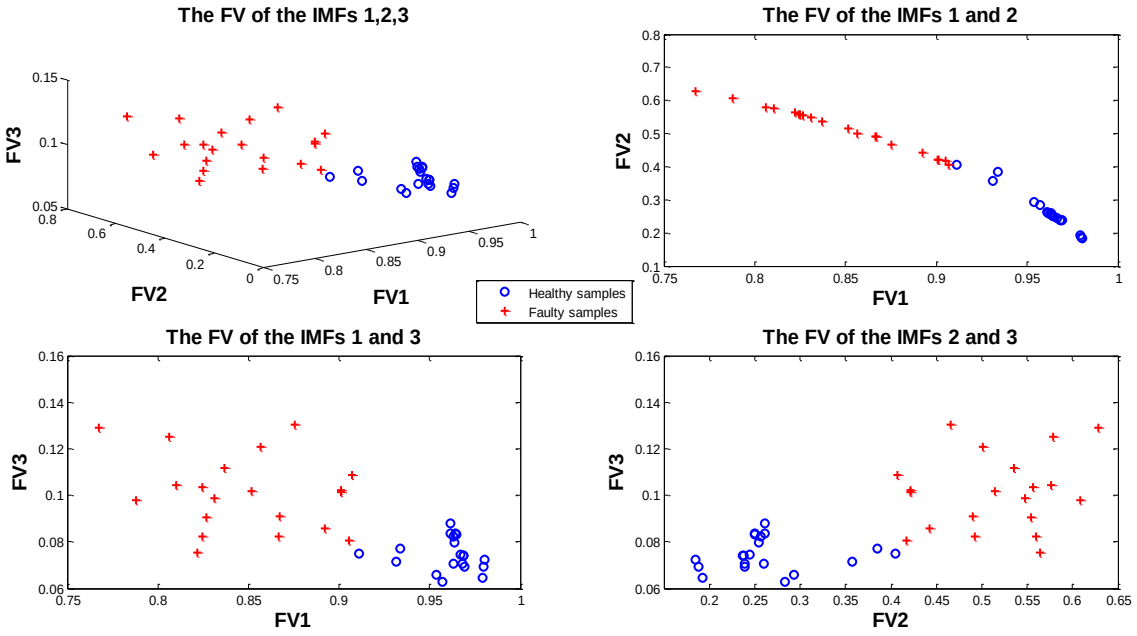


Figure 5.25- The three dimensional feature vectors obtained from the second cointegrating vectors.

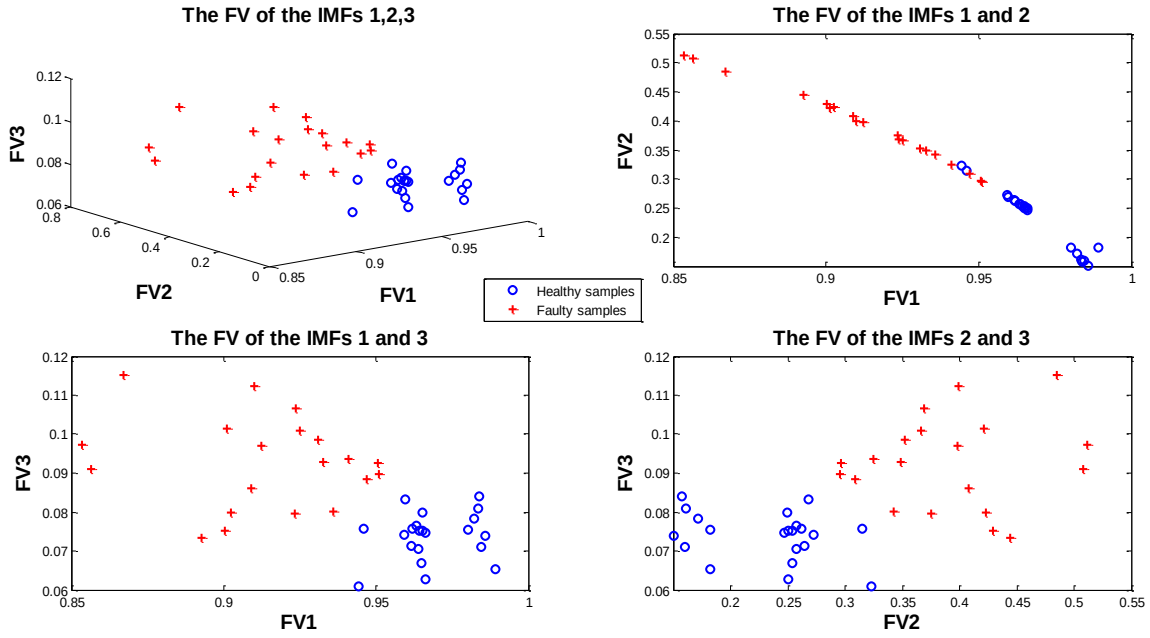


Figure 5.26- The three dimensional feature vectors obtained from the third cointegrating vectors.

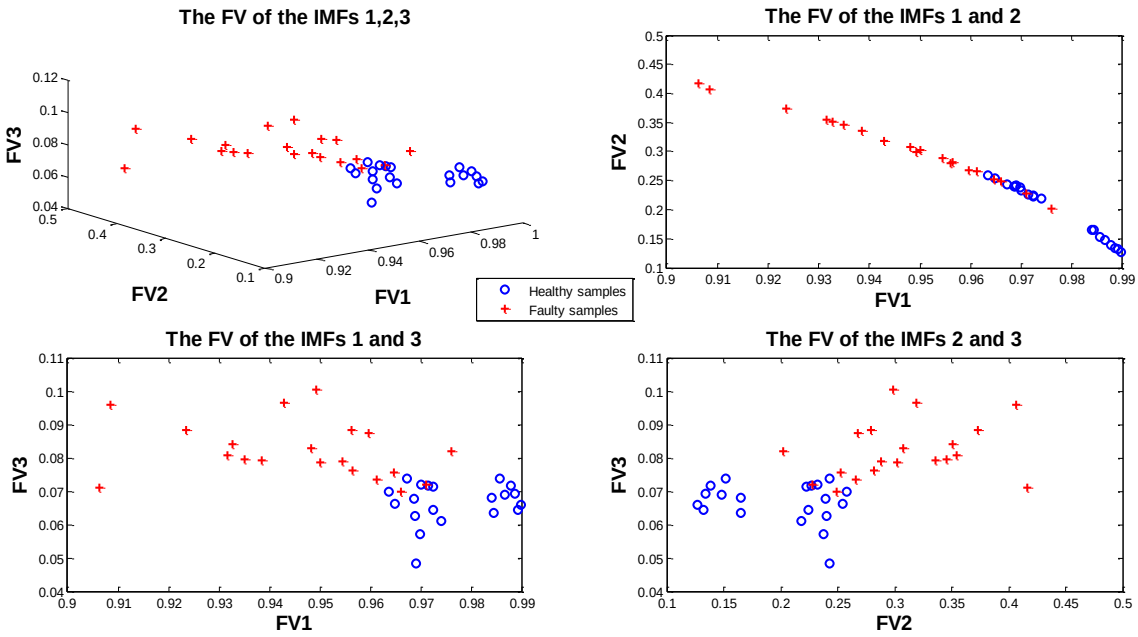


Figure 5.27- The three dimensional feature vectors obtained from the fourth cointegrating vectors.

Table 5.22- The classification results for different cointegrating vectors

Cointegrating vectors	training	test	Success ratio	Margin
1	100	100	24/24	1.00428
2	100	100	24/24	1.00272
3	100	91.7	22/24	0.998041
4	100	83.3	20/24	0.999554

The classification results are evidenced in the Table 5.22. The training and test success rates are 100% for the first and the second cointegrating vectors, as it was expected based on Figs 5.24 and 5.25. Although the training success rate is perfect for the third and the fourth cointegrating vectors, the test samples were not classified perfectly. The third one classified incorrectly two samples, whereas the fourth one labelled wrongly three samples. The Margin obtained in each classification is indicated in Table 5.22. Two first vectors create more reliable hyperplane, as they obtain higher value of margin.

Thus, implementing the method proposed in this study, the state of the bearings could be identified perfectly, although the shaft speed changes significantly.

Conclusions

In this section, a novel combined method based on cointegration is proposed for the development of fault features which are sensitive to the presence of defects while in the same time they are insensitive to changes in the operational conditions. It does not require any additional measurements and can identify defects even for considerable speed variations. The results show that by using the first and the second cointegrating vectors (the higher stationarity level), the normal and damaged sample are completely separable and the classification success rate is 100%.

The proposed method can identify the state of bearings even if the shaft speed changes considerably. Whereas other methods such as the order tracking method can be effectively applied when the speed variation is limited.

Conclusions

In this thesis two main topics were investigated; early damage detection of roller bearings by adopting different techniques and defect detection of roller bearings working in time-varying condition.

For the first part, Empirical mode decomposition (EMD), which has been widely used in recent years for damage detection, was investigated. One of the drawbacks of the EMD is sifting stop criterion, which might affect fault diagnosis results. Influence of the criteria proposed for sifting stop criterion of EMD on early defect detection of roller bearing was deeply investigated along the thesis. The signals collected in various speed and load conditions were decomposed by the EMD using different stopping criteria and the feature vectors obtained from energy content of the IMFs were classified using SVM. The success rate of classification and margin demonstrated that the criterion can affect damage detection results.

Then, a combination method was proposed for early damage detection of roller bearing. The wavelet packet transform (WPT) was applied to the experimental data for denoising and the clean data was break-down into IMFs using Ensemble empirical mode decomposition (EEMD) method. It could successfully distinguish between healthy and faulty bearings.

In order to enhance the EEMD, Performance improved EEMD (PIEEMD) was proposed to determine appropriate value for the amplitude of added noise. The method proposed was validated using the signals collected under various operating conditions.

A novel feature extraction method was also proposed for detection of a small size defect which was only based on healthy bearing signals. The Teager-Kaiser energy operator (TKEO) was applied to IMFs obtained by PIEEMD technique to create new feature vectors as input data for one-class SVM. It was demonstrated that the method can effectively label the new unseen signals and recognize the healthy and faulty bearings.

Finally, a new cointegration based method was proposed for fault diagnosis of roller bearing working in time-varying condition. The signals collected during run-up operating condition were divided into some segments and each segment was decomposed by the PIEEMD. Then, the cointegration method was applied to the IMFs to extract stationary residuals. The feature vectors were created by applying the Teager-Kaiser energy operator (TKEO) to the obtained stationary residuals. Finally, the feature vectors of the healthy bearing signals were utilized as input for one-class support vector machine. The results verified that the method can successfully distinguish between healthy and faulty bearings even if the shaft speed changes dramatically.

Future work that might be of interest would be application of similar signal processing methods to other components of the rotating machines, i.e. gearbox and shaft that might need a completely different approach in the analysis in order to extract damage features. Moreover, the datasets processed in this thesis were collected on the test rig assembled in a laboratory conditions. The methods proposed for damage detection will be validated by applying to data acquired on real industrial rotating machines, where the acquisition and the system noise are by far more relevant. It was demonstrated that the proposed methods could detect early damages. One more improvement that could be made is early damage identification and defect extent recognition.

Further works will investigate the potentiality of the method proposed for damage detection of roller bearings working in time-varying condition. The run-up condition was examined in this study as the time-varying condition. Some other condition with higher speed variations will be checked in order to verify the ability of the method. In addition, the kind of roller bearing defect

examined was limited to a localized defect on inner ring. It would be investigated if the method would be effective for early damage detection and identification.

Bibliography

[1] K. Worden, J.M. Dulieu-Barton, An overview of intelligent fault detection in systems and structures, International Journal of Structural Health Monitoring, 3 (1) (2004), 85-98.

[2] KOYO, JTEKT Corporation, Ball and roller bearings: Failure, causes and countermeasures, Retrieved from JTEKT website: www.jtekt.com.

[3] FAG, Rolling Bearing Damage: Recognition of damage and bearing inspection, Retrieved from website: www.fag.com

[4] TIMKEN, Timken bearing damage analysis with lubrication reference guide, Retrieved from website: www.timken.com.

[5] M. Pirra, Advanced techniques for aircraft bearing diagnostics, PhD thesis, 2012.

[6] P.Y. Kim, A review of rolling element bearing health monitoring (II): preliminary test results on current technologies, In: Proceedings of Machinery Vibration Monitoring and Analysis Meeting, Vibration Institute, New Orleans, LA, 26-28 June, 1984, 127-37.

[7] K. Nishio, S. Hoshiya, T. Miyachi, An investigation of the early detection of defects in ball bearings by the vibration monitoring, ASME Paper 79-DET-45. New York: ASME, 1979.

[8] T. Igarashi, B. Noda, E. Matsushima, A study on the prediction of abnormalities in rolling bearings,1 (1980), J JSLE Int, 71-76.

[9] BT. Kuhnell, JS. Stecki, Correlation of vibration, wear debris analysis and oil analysis in rolling element bearing condition monitoring, Maintenance Management Int, 5 (1985), 105-115.

[10] Z. Reif, MS. Lai, Detection of developing bearing failures by means of vibration, ASME Design Eng Div (Publ) DE, 18 (1) (1989), 231-236.

[11] BVA. Rao, S. Swarnamani, GV. Varghese, Studies on a test rig to check defective and spurious ball and roller bearings, In: Proceedings of the National Conference on Industrial Tribology, Bombay, India, 1986.

[12] J.J. Broderick, RF. Burchill, HL. Clark, Design and fabrication of prototype system for early warning of impending bearing failure, MTI Report MTI-71 TR-1 (prepared for NASA), 1972.

[13] D. Dyer, Bearing condition monitoring. In: Interim Report 1. Southampton (UK): Department of Mechanical Engineering, University of Southampton, 1973.

[14] MS. Darlow, RH. Badgley, Early detection of defects in rolling element bearings, SAE Paper (1975), 750209.

[15] T. Igarashi, H. Hamada, Studies on the vibration and sound of defective rolling bearings (first report: vibration of ball bearings with one defect), Bull JSME, 25(204) (1982), 994-1001.

[16] LG. Martins, SNY Gerges, Comparison between signal analysis for detecting incipient bearing damage, In: Proceedings of the International Condition Monitoring Conference, Swansea, UK, 10-13 April, 1984, 191-204.

[17] AF. Stronach, CJ. Cudworth , AB. Johnston, Condition monitoring of rolling element bearings, In: Proceedings of the International Condition Monitoring Conference, Swansea, UK, 10-13 April, 1984, 162-77.

[18] T. Miyachi, K. Seki, An investigation of the early detection of defects in ball bearings using vibration monitoring- practical limit of detectability and growth speed of defects, In: Proceedings of the International Conference on Rotor-dynamics, Tokyo, 14-17 September, 1986, 403-408.

[19] N. Tandon, A comparison of some vibration parameters for the condition monitoring of rolling element bearings, Measurement, 12 (1994), 285-289.

[20] R. Prabhu R, Rolling bearing diagnostics. In: Proceedings of the Indo-US Symposium on Emerging Trends in Vibration and Noise Engineering, New Delhi, 18-20 March, 1996, 311-320.

[21] N.V. Kirianaki, S.Y. Yurish, N.O. Shpak, V.P. Deynega, Data Acquisition and Signal Processing for Smart Sensors, Wiley, Chichester, West Sussex, England, 2002.

[22] A. K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process., 20 (2006), 1483-1510.

[23] J. S. Mitchell, Introduction to Machinery Analysis and Monitoring, PenWel Books, Tulsa, 1993.

[24] A. R. Crawford, The Simplified Handbook of Vibration Analysis, Computational Systems, Inc., Knoxville, 1992.

[25] V. Wouk, Machinery Vibration Measurement and Analysis, McGraw-Hill, New York, 1991.

[26] R.C. Eisenmann, Sr. and R.C. Eisenmann, Jr., Machinery Malfunction Diagnosis and Correction: Vibration Analysis and Troubleshooting for the Process Industries, Hewlett-Packard Professional books, Prentice-Hall, Upper Saddle River, NJ, 1997.

[27] Effective Machinery Measurements using Dynamic Signal Analysers, Application Note 243-1, Hewlett Packard Company, 1997.

[28] J. I. Taylor, Back to the Basics of Rotating Machinery Vibration Analysis, Vibration Consultants, Inc., Tampa Bay, FL, 1994.

[29] K. Worden, W. J. Staszewski, J. J. Hensman, Natural computing for mechanical systems research: A tutorial overview, Mech. Syst. Signal Process., 25 (2011), 4-111

[30] R. B. Randall, J. Antoni, Rolling element bearing diagnostics – A tutorial, Mech. Syst. Signal Process., 25 (2011), 485-520.

[31] H.L. Balderston, The detection of incipient failure in bearings, Material Evaluation, 27 (1969), 121-128.

[32] B. Weichbrodt, K.A. Smith, Signature analysis - non-intrusive techniques for incipient failure identification, Technical report, General Electric Technical Information Series 70-C-364, 1970.

[33] S. Braun, The extraction of periodic waveforms by time domain averaging, Acoustica, 23 (2) (1975), 69-77.

[34] S. Braun, B. Datner, Analysis of roller/ball bearings, Journal of Design, (101) (1) (1979), 118-128.

[35] M.S. Darlow, R.H. Badgley, G.W. Hogg, Application of high frequency resonance techniques for bearing diagnostics in helicopter gearboxes, Technical Report, US army air mobility research and development laboratory, 1974, 74-77.

[36] J. Mathew, R.J. Alfredson, The condition monitoring of rolling element bearings using vibration analysis, Trans ASME, J Vibr, Acoust, 106 (1984), 447-453.

[37] P.D. McFadden, J.D. Smith, Vibration monitoring of rolling element bearings by the high frequency resonance technique - a review, Tribol Int, 17 (1) (1984), 3-10.

[38] N. Tandon, B.C. Nakra, Vibration and acoustic monitoring techniques for the detection of defects in rolling element bearings - a review, Shock Vibr Digest, 24 (3) (1992), 3-11.

[39] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearing, Tribology International, 32 (1999), 469-480.

[40] I.M. Howard, A review of rolling element bearing vibration: detection, diagnosis and prognosis, Technical report, DSTO Research report 0013, 1994.

[41] R.B. Randall, State of the art in monitoring rotating machinery - part 1, Sound and Vibration, 38 (3) (2004),14-21.

[42] R.B. Randall, State of the art in monitoring rotating machinery - part 2, Sound and Vibration, 38 (5) (2004),10-17.

[43] C. Pachaud, R. Salvetat, C. Fray, Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces, Mech. Syst. Signal Process., 11 (1997), 903-916.

[44] J.P. Dron, F. Bolaers, I. Rasolofondraibe, Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearings, J. Sound Vib., (2004) 61-73270 (2004), 61-73.

[45] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control, fourth ed., Wiley-Blackwell, 2008.

[46] H. Sohn, C.R. Farrar, Damage diagnosis using time series analysis of vibration signals, Smart Materials and Structures 10 (2001), 446-451.

[47] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Cambridge University Press, 1992.

[48] T. Liu, J. Chen, C. Dong,W. Xiao, and X. Zhou, The fault detection and diagnosis in rolling element bearings using frequency band entropy. In: Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2012.

[49] B. P. Bogert, M. J. R. Healy, J. W. Tukey, The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking, In: Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15, New York: Wiley, 1963, 209-243.

[50] M. Inalpolat, A. Kahraman, A theoretical and experimental investigation of modulation sidebands of planetary gear sets. J. Sound Vib., 323(2009), 677-696.

[51] M. Inalpolat, A. Kahraman, A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors, J. Sound Vib., 329(4) (2010), 371-393.

[52] M. El Badaoui, F. Guillet, J. Danière, New applications of the real cepstrum to gear signals, including definition of a robust fault indicator, Mech. Syst. Signal Process., 18(5)(2004), 1031-1046.

[53] Z. Feng, M. Liang, F. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples, Mech. Syst. Signal Process., 38 (2013), 165-205.

[54] F. Hlawatsch, G.F. Boudreaux-Bartels, Linear and quadratic time-frequency signal representations, IEEE Signal Process., 9 (2) (1992), 21-67.

[55] W.J. Wang, P.D. McFadden, Early detection of gear failure by time-frequency analysis. Part 1: Calculation of the time-frequency distribution, Mech. Syst. Signal Process., 7 (1993), 193-204.

[56] J. Antoni, The spectral kurtosis: a useful tool for characterizing nonstationary signals, Mech. Syst. Signal Process., 20(2) (2006), 282-307.

[57] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process., 20 (2) (2006), 308-331.

[58] Ch.K. Chui, Wavelet Analysis and its Applications. I: An Introduction to Wavelets, Academic Press, 1992.

[59] Ch.K. Chui, Wavelet Analysis and its Applications. II: Wavelets: A Tutorial in Theory and Applications, Academic Press, 1992.

[60] Ch.K. Chui, Wavelets: A Mathematical Tool for Signal Processing, SIAM, 1997

[61] Z. Peng, F. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mech. Syst. Signal Process., 18(2004), 199-221.

[62] R. Yan, R.X Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing, 96 (2014), 1-15.

[63] K.R. Fyfe, E.D.S. Munck, Analysis of computed order tracking, Mech. Syst. Signal Process., 11 (1997), 189-205.

[64] P. Borghesani, R. Ricci, S. Chatterton, P. Pennacchi, A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions, Mech. Syst. Signal Process., 38 (2013), 23-35.

[65] P. Borghesani, P. Pennacchi, S. Chatterton, R. Ricci, The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery, Mech. Syst. Signal Process., 44 (2014), 118-133.

[66] M.D. Coats, R.B. Randall, Single and multi-stage phase demodulation based order-tracking, Mech. Syst. Signal Process., 44 (2014), 86-117.

[67] H. Andre, Z. Daher, J. Antoni, D. Rémond, Comparison between angular sampling and angular resampling methods applied on the vibration monitoring of a gear meshing in non-stationary conditions, In: Proceedings of ISMA, 2010.

[68] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, D. Remond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, Mech. Syst. Signal Process., 24 (2010), 1998-2011.

[69] L.F. Villa, A. Reñones, J.R. Perán, L.J. de Miguel, Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation, Mech. Syst. Signal Process., 25 (2011), 2157-2168.

[70] E.P. Wigner, On the quantum correction for thermodynamic equilibrium. Physical Review, 40 (1932), 749-759.

[71] J. Ville, Théorie et applications de la notion de signal analytique, Cables et Transmission, 2 A (1948), 61-74.

[72] B.D. Forrester, Analysis of gear vibration in the time-frequency domain, In: Proceedings of the 44th Meeting of the Mechanical Failures Group of the Vibration Institute, Virginia Beach, 1990, 225-239.

[73] B.D. Forrester, Analysis time-frequency analysis in machine fault detection, In: B. Boashash (Ed.), Time-Frequency Signal, Longman Cheshire, 1992.

[74] W.J. Staszewski, G.R. Tomlinson, Time-variant methods in machinery diagnostics, In: H.G. Natke, G.R. Tomlinson, J.T.P. Yao (Eds.), Safety Evaluation Based on Identification Approaches Related to Time-Variant and Nonlinear Structures, Vieweg, 1993.

[75] P.D. McFadden, W. Wang, Time-frequency domain analysis of vibration signals for machinery diagnostics: (1) introduction to the Wigner-Ville distribution, Research Report OUEL 1859/90, Department of Engineering, Oxford University, 1990.

[76] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H.Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London Series A 454 (1998), 903-995.

[77] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process., 19 (2) (2005) 259-270.

[78] J. Cheng, D. Yu, J.S. Tang, Y. Yang, Application of frequency family separation method based upon EMD and local Hilbert energy spectrum method to gear fault diagnosis, Mech. Mach. Theory, 43 (6) (2008), 712-723.

[79] J. Cheng, D. Yu, J. Tang, Y. Yang, Local rub-impact fault diagnosis of the rotor systems based on EMD, Mech. Mach. Theory, 44 (4) (2009), 784-791.

[80] Z. Peng, P.W. Tse, F. Chu, A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing, Mech. Syst. Signal Process., 19 (2005), 974–988.

[81] Z. Peng, P.W. Tse, F. Chu, An improved Hilbert-Huang transform and its application in vibration signal analysis, J. Sound Vib., 286 (1-2) (2005), 187-205.

[82] D. Guo, Z. Peng, Vibration analysis of a cracked rotor using Hilbert–Huang transform, Mech. Syst. Signal Process., 21 (8) (2007), 3030-3041.

[83] W. Yang, Interpretation of mechanical signals using an improved Hilbert–Huang transform, Mech. Syst. Signal Process., 22 (5) (2008), 1061-1071.

[84] W. Yang, P.J. Tavner, Empirical mode decomposition, an adaptive approach for interpreting shaft vibratory signals of large rotating machinery, J. Sound Vib., 321 (3-5) (2009), 1144-1170.

[85] F. Wu, L. Qu, An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery, J. Sound Vib., 14 (3–5) (2008), 586–602.

[86] F. Wu, L. Qu, Diagnosis of sub-harmonic faults of large rotating machinery based on EMD, Mech. Syst. Signal Process., 23 (2) (2009), 467-475.

[87] Q. Gao, C. Duan, H. Fan, Q. Meng, Rotating machine fault diagnosis using empirical mode decomposition, Mech. Syst. Signal Process., 22 (5) (2008), 1072-1081.

[88] G. Gai, The processing of rotor start-up signals based on empirical mode decomposition, Mech. Syst. Signal Process., 20 (1) (2006), 222-235.

[89] T. Ramesh Babu, S. Srikanth, A.S. Sekhar, Hilbert-Huang transform for detection and monitoring of crack in a transient rotor, Mech. Syst. Signal Process., 22 (4) (2008), 905-914.

[90] S.J. Loutridis, Damage detection in gear systems using empirical mode decomposition, Eng. Struct, 26 (2004), 1833-1841.

[91] B. Liu, S. Riemenschneider, Y. Xu, Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum, Mech. Syst. Signal Process., 20 (3) (2006), 718-734.

[92] Y. Lei, Z. He, Y. Zi, Application of the EEMD method to rotor fault diagnosis of rotating machinery, Mech. Syst. Signal Process., 23 (4) (2009), 1327-1338.

[93] Y. Lei, M.J. Zuo, Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs, Meas. Sci. Technol., 20 (12) (2009), 1-12.

[94] P. Maragos, J.F. Kaiser, T.F. Quatieri, On amplitude and frequency demodulation using energy operators, IEEE Trans. Signal Process., 41 (4) (1993), 1532–1550.

[95] P. Maragos, J.F. Kaiser, T.F. Quatieri, Energy separation in signal modulations with application to speech analysis, IEEE Trans. Signal Process., 41 (10) (1993), 3024-3051.

[96] C. Bovik, P. Maragos, T.F. Quatieri, AM-FM energy detection and separation in noise using multiband energy operators, IEEE Trans. Signal Process., 41 (12) (1993), 3245–3265.

[97] A. Potamianos, P. Maragos, A comparison of the energy operator and Hilbert transform approaches for signal and speech demodulation, Signal Process, 37 (1) (1994), 95–120.

[98] J. Cheng, D. Yu, Y. Yang, The application of energy operator demodulation approach based on EMD in machinery fault diagnosis, Mech. Syst. Signal Process., 21 (2) (2007), 668–677.

[99] M. Bassiuny, X. Li, Flute breakage detection during end milling using Hilbert-Huang transform and smoothed nonlinear energy operator, Int. J. Mach. Tools Manuf., 47 (6) (2007), 1011–1020.

[100] H. Li, H. Zheng, L. Tang, Gear fault detection based on Teager-Huang transform, Int. J. Rotating Mach., 2010 (2010), 502064.

[101] M. Liang, I. Soltani Bozchalooi, An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection, Mech. Syst. Signal Process., 24 (5) (2010), 1473-1494.

[102] I. Soltani Bozchalooi, M. Liang, Teager energy operator for multi-modulation extraction and its application for gearbox fault detection, Smart Mater. Struct., 19 (2010), 075008.

[103] H. Li, H. Zheng, L. Tang, Bearing fault detection and diagnosis based on Teager-Huang transform, Int. J. wavelets multiresolution and information Processing, 7 (5) (2009), 643-663.

[104] Z. Feng, T. Wang, M. Zuo, F. Chu, S. Yan, Teager energy spectrum for fault diagnosis of rolling element bearings, J. Phys., Confer. Ser., 305(2011), 012129.

[105] H. Liu, J. Wang, C. Lu, Rolling bearing fault detection based on the Teager energy operator and Elman neural network, Mathematical problems in engineering, 10 (2013), Article ID 498385.

[106] P. Rodriguez, J. Alonso, M. Ferrer, C. Travieso, Application of the Teager-Kaiser energy operator in bearing fault diagnosis, ISA Transactions, 52 (2013), 278-284.

[107] K. Worden, G. Manson, N.R.J. Fieller, Damage detection using outlier analysis, J. Sound Vib., 229 (2000), 647-667.

[108] L. Tarassenko, P. Hayton, Z. Cerneaz, M. Brady, Novelty detection for the identification of masses in mammograms. Proceedings of 4th International Conference on Neural Networks, Cambridge, UK. IEE Publication 409 (1995), 442-447.

[109] D.A. Pomerleau, Input reconstruction reliability information. In: Hanson, S.J., Cowan, J.D. and Giles, C.L. (eds.), Advances in neural information processing systems 5 (1993), Morgan Kaufman Publishers.

[110] O. Taylor, J. MacIntyre, C. Isbell, C. Kirkham, A. Long, Adaptive fusion devices for condition monitoring: local fusion systems of the NEURALMAINE project, In: Proceedings of 1st International Conference on Damage Assessment of Structures, DAMAS 1999, 205-216.

[111] S. Roberts, L. Tarassenko, A probabilistic resource allocating network for novelty detection. Neural Computation, 6 (1994), 270-284.

[112] H. Sohn, C.R. Farrar, Statistical process control and projection techniques for damage detection. In: Proceedings of European COST F3 Conference on System Identification and Structural Health Monitoring, Madrid, Spain, 2000, 105-114.

[113] B. Scholkopf, R. Williamson, A. Smola, J. S. Taylor, J. Platt, Support vector method for novelty detection, Advances in Neural Information Processing Systems, 12 (2000), 582-586.

[114] W.J. Staszewski, K. Worden, Classification of faults in gearboxes pre-processing algorithms and neural networks. Neural Computing and Applications, 5(3) (1997), 160-183.

[115] J. Rafiee, F. Arvani, A. Hari, M.H. Sadeghi, Intelligent condition monitoring of a gearbox using artificial neural network, Mech. Syst. Signal Process., 21(4) (2007), 1746-1754.

[116] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process., 21(6) (2007), 2560 -2574.

[117] J.J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, New York, 1998.

[118] S. Simani, C. Fantuzzi, R.J. Patton, Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques, Springer, London, 2003.

[119] M. Nyberg, Model based fault diagnosis methods, theory, and automotive engine applications, PhD thesis, 1999.

[120] I. Howard, S. Jia, J. Wang, The dynamic modelling of a spur gear in mesh including friction and a crack, Mech. Syst. Signal Process., 15 (2001), 831-838.

[121] W.Y. Wang, Towards dynamic model-based prognostics for transmission gears, In: Component and Systems Diagnostics, Prognostics, and Health Management II, vol. 4733, Bellingham, 2002, 157-167.

[122] N. Sawalhi, R.B. Randall, Simulating gear and bearing interactions in the presence of fault, Part 1 : The combined gear bearing dynamic model and the simulation of localized bearing faults, Mech. Syst. Signal Process., 22 (2008), 1924-1951.

[123] D.C. Baillie, J. Mathew, Nonlinear model-based fault diagnosis of bearings, In: Proceedings of an International Conference on Condition Monitoring, Swansea, UK, 1994, 241-252.

[124] K.A. Loparo, M.L. Adams, W. Lin, M.F. Abdel-Magied, N. Afshari, Fault detection and diagnosis of rotating machinery, IEEE Transactions on Industrial Electronics, 47 (2000), 1005-1014.

[125] K.A. Loparo, A.H. Falah, M.L. Adams, Model-based fault detection and diagnosis in rotating machinery, In: Proceedings of the Tenth International Congress on Sound and Vibration, Stockholm, Sweden, 2003, 1299-1306.

[126] A. Rafsanjani, S. Abbasion, A. Farshidianfar, H. Moeenfard, Nonlinear dynamic modelling of surface defects in rolling element bearing systems, J. Sound Vib., 319 (2009), 1150-1174.

[127] C.H. Oppenheimer, K.A. Loparo, Physically based diagnosis and prognosis of cracked rotor shafts, In: Component and Systems Diagnostics, Prognostics, and Health Management II, 4733, Bellingham, 2002, 122-132.

[128] A.S. Sekhar, Model-based identification of two cracks in a rotor system, Mech. Syst. Signal Process., 18 (2004), 977-983.

[129] R.E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering, 82 (1) (1960), 35-45.

[130] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[131] J. Wu, C. Huang, R. Huang, An application of a recursive Kalman filtering algorithm in rotating machinery fault diagnosis, NDT&E International, 37 (3) (2004), 411-419.

[132] K. Li, Y. Zhang, Z. Li, Application research of Kalman filter and SVM applied to condition monitoring and fault diagnosis, Appl. Mech. Mat., 121-126 (2012), 268-272.

[133] S. Khanam, J. K. Dutt, N. Tandon, Extracting rolling element bearing faults from noisy vibration signal using kalman filter, J. Vib. Acoust., 136 (3) (2014) (11 pages), doi: 10.1115/1.4026946.

[134] B. Widrow, S. Stearns, adaptive signal processing, Englewood Clis NJ: Prentice-Hall, 1985.

[135] G.K. Chaturvedi, D.W. Thomas, Bearing fault detection using adaptive noise cancelling, J. Sound Vib., 104 (1982), 280-289.

[136] C.C. Tan, B. Dawson, An adaptive noise cancellation approach for condition monitoring of gearbox bearings, In: Proceedings of the international tribology conference, Melbourne, 1987.

[137] D. Ho, Bearing diagnostics and self-adaptive noise cancellation, Ph.D. Thesis, University of

New South Wales, Australia, 1990.

[138] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals. Part I—evaluation of adaptive algorithms, Mech. Syst. Signal Process., 18 (1) (2003), 89-101.

[139] D. Ho, R.B. Randall, Effects of time delay, order of fir filter and convergence factor on self adaptive noise cancellation, ICSV5, Adelaide, 1997.

[140] R.B. Randall, Y. Li, Diagnostics of planetary gear bearings in the presence of gear vibrations, Proceedings of second international conference on gearbox vibration and diagnostics, Imeche, London, (1995), 73-80.

[141] S.K. Lee, P.R. White, The enhancement of impulsive noise and vibration signals for fault detection in rotating and reciprocating machinery, J. Sound Vib., 217 (3) (1998), 485-505.

[142] T. Barszcz, Decomposition of vibration signals into deterministic and nondeterministic components and its capabilities of fault detection and identification, Int. J. Appl. Math. Comput. Sci., 19 (2) (2009), 327-335.

[143] J. Antoni, R.B. Randall, Optimization of SANC for Separating Gear and Bearing Signals, Comadem'01, Manchester, UK, 2001.

[144] S. Mallat, A theory of multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (7) (1989), 674-693.

[145] MATLAB 6.5, Release 13, Wavelet Toolbox, The Math Works, Inc., 2002.

[146] F. Al-Badour, M. Sunar, and L. Cheded, Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques, Mech. Syst. Signal Process., 25 (2011), 2083-2101.

[147] J. Rafiee, P.W. Tse, Use of autocorrelation of wavelet coefficients for fault diagnosis, Mech. Syst. Signal Process., 23 (5) (2009), 1554-1572.

[148] J. Rafiee, P.W. Tse, A. Harifi, M.H. Sadeghi, A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system, Expert Syst. Appl., 36 (3) (2009), 4862-4875.

[149] J. Rafiee, M.A. Rafiee, P.W. Tse, Application of mother wavelet functions for automatic gear and bearing fault diagnosis, , Expert Syst. Appl., 37 (6) (2010), 4568-4579.

[150] R. Coifman, M. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Transactions on information theory, 38 (1992), 713-718.

[151] D.L. Donoho, I.M. Johnstone, Ideal denoising in an orthonormal basis chosen from a library of bases, Comptes Rendus De L Academie Des Sciences Serie I-Mathematique, 319 (12), (1994), 1317-1322.

[152] C.M. Stein, Estimation of the mean of a multivariate normal-distribution, Annals of statistics 9 (6) (1981), 1317-1322.

[153] J. Lin, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, J. Sound Vib., 234 (2000), 135-148.

[154] D. Donoho, De-noising by soft thresholding. IEEE Transactions on information theory, 41 (1995), 613-627.

[155] Z. Peng, F. Chu, Application of the wavelet transform in machine condition monitoring and

fault diagnostics: A review with bibliography. Mech. Syst. Signal Process., 18 (2004), 199-221.

[156] R. Yan, R.X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing, 96 (2014), 1-15.

[157] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, J. Sound Vib., 289 (2006), 1066-1090.

[158] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, Mech. Syst. Signal Process., 21 (7) (2007), 2933-2945.

[159] M.A. Jafarizadeh, R. Hassannejad, M.M. Ettefagh, S. Chitsaz, Asynchronous input gear damage diagnosis using time averaging and wavelet filtering, Mech. Syst. Signal Process., 22 (2008), 172-201.

[160] K.F. Al-Raheem, A. Roy, K.P. Ramachandran, D.K. Harrison, S. Grainger, Rolling element bearing faults diagnosis based on autocorrelation of optimized wavelet denoising technique, International journal of advanced manufacturing technology, 40 (2009), 393-402.

[161] W. He, Z. Jiang, K. Feng, Bearing fault detection based on optimal wavelet filter and sparse code shrinkage, Measurement, 42 (2009), 1092-1102.

[162] W. Su, F. Wang, H. Zhu, Z. Zhang, Z. Guo, Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement, Mech. Syst. Signal Process., 24 (2010), 1458-1472.

[163] X. Chiementin, D. Mba, B. Charnley, S. Lignon, J.P. Dron, Effect of the denoising on Acoustic Emission signals, J. Vib. Acoust., 132 (3) (2010), 0310091 1-9.

[164] Z. Li, X.Yan, C. Yuan, Z. Peng, L. Li, Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method, Mech. Syst. Signal Process., 25 (2011), 2589-2607.

[165] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnostics, Mech. Syst. Signal Process., 15 (2001), 963-977.

[166] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing, Mech. Syst. Signal Process., 19 (2005), 974-988.

[167] Y. Lei, J. Lin, Z. He, M. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mech. Syst. Signal Process., 35 (2013), 108-126.

[168] P. Flandrin, P. Gon¸calv`es and G. Rilling, EMD equivalent filter banks, from interpretation to applications, in Hilbert–Huang Transform: Introduction and Applications, eds. N. E. Huang and S. S. P. Shen, World Scientific, Singapore, 2005, 67-87.

[169] Z. Wu, N.E. Huang, A study of the characteristics of white noise using the Empirical Mode Decomposition method, Proc. Roy. Soc. London A, 2004.

[170] Q. Gao, C. Duan, H. Fan, and Q. Meng, Rotating machine fault diagnosis using empirical mode decomposition, Mech. Syst. Signal Process., 22 (2008), 1072-1081.

[171] Z. Wu, N. Huang, Ensemble Empirical Mode Decomposition: A noise-assisted data analysis

method, Advances in adaptive data analysis, 1 (1) (2009), 1-41.

[172] P. Flandrin, G. Rilling, P. Gonc-alve´ s, Empirical mode decomposition as a filter bank, IEEE Signal Process., 11 (2004), 112-114.

[173] Q.H. Du, S.N. Yang, Improvement of the EMD method and applications in defect diagnosis of ball bearings, Meas. Sci. Technol., 17 (2006), 2355-2361.

[174] Q.H. Du, S.N. Yang, Application of the EMD method in the vibration analysis of ball bearings, Mech. Syst. Signal Process., 21 (2007), 2634-2644.

[175] H.B. Dong, K.Y. Qi, X.F. Chen, et al., Sifting process of EMD and its application in rolling element bearing fault diagnosis, J. Mech. Sci. Technol., 23 (2009), 2000-2007.

[176] J. Terrien, C. Marque, B. Karlsson, Automatic detection of mode mixing in empirical mode decomposition using non-stationarity detection: application to selecting IMFs of interest and denoising, EURASIP J. Adv. Signal Process., (2011) 2011, 1-8.

[177] R.Q. Yan, R.X. Gao, Rotary machine health diagnosis based on empirical mode decomposition, J. Vib. Acoust., 130 (2008), 021007.

[178] D.J. Yu, J.S. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process., 19 (2005), 259-270.

[179] F. Chen, X. Zhou, Q.H. Wu, et al., Application of Hilbert–Huang Transformation to fault diagnosis of rotary machinery, Fifth International Symposium on Instrumentation Science and Technology, Shenyang, China, September 15-18, 2008.

[180] H. Li, H.Q. Zheng, Bearing fault detection using envelope spectrum based on EMD and TKEO, Proceedings of the Fifth international conference on fuzzy systems and knowledge discovery, Shandong, China, October 18-20, 2008, 142-146.

[181] V.K. Rai, A.R. Mohanty, Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform, Mech. Syst. Signal Process., 21 (2007), 2607-2615.

[182] H. Li, H.Q. Zheng, L.W. Tang, Wigner–Ville distribution based on EMD for faults diagnosis of bearing, Fuzzy Syst. Knowl. Discovery, 4223 (2006), 803-812.

[183] H. Li, Y.P. Zhang, H.Q. Zheng, Bearing fault detection and diagnosis based on order tracking and Teager–Huang transform, J. Mech. Sci. Technol., 24 (2010), 811-822.

[184] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing, Mech. Syst. Signal Process., 19 (2005), 974-988.

[185] Y. Yang, D.J. Yu, J.S. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, Measurement, 40 (2007), 943-950.

[186] J.S. Cheng, D.J. Yu, Y. Yang, A fault diagnosis approach for roller bearings based on EMD method and AR model, Mech. Syst. Signal Process., 20 (2006), 350-362.

[187] Y. Yang, D.J. Yu, J.S. Cheng, A roller bearing fault diagnosis method based on EMD energy entropy and ANN, J. Sound Vib., 294 (2006), 269-277.

[188] J.S. Cheng, D.J. Yu, J.S. Tang, et al., Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery, Shock Vib., 16 (2009), 89-98.

[189] Y.G. Lei, Z.J. He, Y.Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Syst. Appl., 35 (2008), 1593-1600.

[190] Y.G. Lei, Z.J. He, Y.Y. Zi, Application of a novel hybrid intelligent method to compound fault diagnosis of locomotive roller bearings, J. Vib. Acoust., 130 (2008), 034501.

[191] Y.G. Lei, Z.J. He, Y.Y. Zi, et al., Fault diagnosis of rotating machinery based on multiple ANFIS combination with Gas, Mech. Syst. Signal Process., 21 (2007), 2280-2294.

[192] X.L. An, D.X. Jiang, S.H. Li, et al., Application of the ensemble empirical mode decomposition and Hilbert transform to pedestal looseness study of direct-drive wind turbine, Energy, 36 (2011), 5508–5520.

[193] S.F. Ai, H. Li, Y.P. Zhang, Condition monitoring for bearing using envelope spectrum of EEMD, International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, China, April 11-12, 2009, 190-193.

[194] M. Zvokelj, S. Zupan, I. Prebil, Non-linear multivariate and multiscale monitoring and signal denoising strategy using kernel principal component analysis combined with ensemble empirical mode decomposition method, Mech. Syst. Signal Process., 25 (2011), 2631-2653.

[195] J. Zhang, R.Q. Yan, R.X. Gao, et al., Performance enhancement of ensemble empirical mode decomposition, Mech. Syst. Signal Process., 24 (2010), 2104-2123.

[196] X.M. Lu, J. Wang, Bearing fault diagnosis based on redundant second generation wavelet denoising and EEMD, International Conference on Consumer Electronics, Communications and Networks, XianNing, China, April 16-18, 2011, 1090-1093.

[197] Y.G. Lei, Z.J. He, Y.Y. Zi, EEMD method and WNN for fault diagnosis of locomotive roller bearings, Expert Syst. Appl., 38 (2011), 7334-7341.

[198] W. Guo, P.W. Tse, Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis, 2010 Prognostics & System Health Management Conference, Macao, January 12–14, 2010, 1-7.

[199] H. M. Teager, Some observations on oral flow during phonation, IEEE Trans. Acoustics, Speech and Signal Process., 28 (5) (1980), 599-601.

[200] J. F. Kaiser, On a simple algorithm to calculate the ‘energy’ of a signal, In Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process., Albuquerque, NM, 1990, 381-384.

[201] R. Hamila, J. Astola, F. Alaya Cheikh, M. Gabbouj, M. Renfors, Teager energy and the ambiguity function, IEEE Transactions on Signal Processing, 47 (1999), 260-262.

[202] E. Kvedalen, Signal processing using the Teager energy operator and other nonlinear operators, PhD thesis, 2003.

[203] C. Junsheng, Y. Dejie, Y. Yu, The application of energy operator demodulation approach based on EMD in machinery fault diagnosis, Mech. Syst. Signal Process., 21 (2007), 668-677.

[204] H. Li, L. Fu, Y. Zhang, Bearing fault diagnosis based on Teager energy operator demodulation technique, In: Measuring technology and mechatronics automation, April 11-12, 2009.

[205] H. Li, H. Zheng, L. Tang, Bearing fault detection and diagnosis based on Teager-Huang transform, Int. J. wavelets multiresolution and information Processing, 7 (5) (2009), 643-663.

[206] Z. Feng, T. Wang, M. Zuo, F. Chu, S. Yan, Teager energy spectrum for fault diagnosis of rolling element bearings, J. Phys.: Confer. Ser., 305 (2011), 012129.

[207] H. Liu, J. Wang, C. Lu, Rolling bearing fault detection based on the Teager energy operator and Elman neural network, Mathematical problems in engineering, 10 (2013), Article ID 498385.

[208] P. Rodriguez, J. Alonso, M. Ferrer, C. Travieso, Application of the Teager-Kaiser energy operator in bearing fault diagnosis, ISA Transactions, 52 (2013), 278-284.

[209] D. Kwak, D. Lee, J. Ahn, B. Koh, Fault detection of roller-bearings using signal processing and optimization algorithms, Sensors, SA Transactions, 14(1) (2014), 283-298.

[210] R. Engle, C. Granger. Co-integration and error-correction: representation, estimation, and testing. Econometrica, 55 (1987), 251-276.

[211] S. Johansen, Likelihood-based inference in cointegrated vector autoregressive models, Oxford: Oxford university press, 1995.

[212] D.A. Dickey, W.A. Fuller: Journal of the American Statistical Association, 75 (1979), 427-431.

[213] J. MacKinnon, Numerical Distribution Functions for Unit Root and Cointegration Tests, Journal of Applied Econometrics, 11 (1996), 601-618.

[214] T. Mills, The Econometric Analysis of Financial Time Series, Cambridge University Press, Cambridge, 1999.

[215] C. Alexander, Market Models: A Guide to Financial Data Analysis, John Wiley and Sons, 2001.

[216] J. Cochrane, Asset Pricing. Princeton University Press, New Jersey, 2001.

[217] R. Tsay, The Analysis of Financial Time Series, John Wiley & Sons, New York, 2001.

[218] E.J. Cross, K. Worden, Cointegration and why it works for SHM, Journal of Physics: Conference Series, 382 (2012), 012046.

[219] E.J. Cross, K. Worden, and Q. Chen. Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 467(2133) (2011), 2712-2732.

[220] I. Antoniadou, E.J. Cross, K. Worden, Cointegration and the empirical mode decomposition for the analysis of the diagnostic data, Key engineering materials, 569-570 (2013), 884-891.

[221] K. Worden, E.J. Cross, I. Antoniadou, A. Kyprianou, A multi resolution approach to cointegration for enhanced SHM of structures under varying conditions- An exploratory study, Mech. Syst. Signal Process., 47 (1-2) (2014), 243-262.

[222] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education , 2005.

[223] D. Coomans, D.L. Massart, Alternative k-nearest neighbor rules in supervised pattern recognition: Part 1. K-Nearest neighbour classification by using alternative voting rules, Analytica Chimica Acta 136 (1982), 15-27.

[224] B.S. Everitt, S. Landau, M. Leese, D. Stahl, Miscellaneous Clustering Methods, in Cluster Analysis, 5th Edition, John Wiley & Sons, Ltd, Chichester, UK, 2011.

[225] C. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearing, Part II: The use of nearest neighbour classification, Mech. Syst. Signal Process., 6 (1992), 309-316.

[226] L. Zhang, L. B. Jack, A. K. Nandi, Extending genetic programming for multi-class classification by combining. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2005.

[227] I. Trendafilova, An automated procedure for detection and identification of ball bearing damage using multivariate statistics and pattern recognition, Mech. Syst. Signal Process., 24 (6) (2010), 1858-1869.

[228] D. He, R. Li, J. Zhu, M. Zade, A data mining based full ceramic bearing fault diagnostic system using ae sensors, IEEE Transactions on Neural Networks 22 (12) (2011), 2022-2031.

[229] L. Jiang, J. Xuan, T. Shi, Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis, Mech. Syst. Signal Process., (41) (1-2) 2013, 113-126.

[230] L. Jiang, T. Shi1, J. Xuan, Fault diagnosis of rolling bearings based on Marginal Fisher analysis, Journal of Vibration and Control 20 (3) (2014), 470-480.

[231] C.M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006.

[232] H. Yang, J. Mathew, L. Ma, Intelligent diagnosis of rotating machinery faults - A review. In: 3rd Asia-Pacific Conference on Systems Integrity and Maintenance, ACSIM, September 2002, Cairns, Australia.

[233] A.C. McCormick, A.K. Nandi, A comparison of artificial neural networks and other statistical methods for rotating machine condition classification, IEE Colloquium on Modelling and Signal Processing for Fault Diagnosis (1996) (Ref. No.1996/260).

[234] B. Li, M.-Y. Chow, Y. Tipsuwan, J. C. Hung, Neural-network-based motor rolling bearing fault diagnosis. IEEE Transactions on industrial electronics 47 (5) (2000), 1060-1069.

[235] B. Samanta, K. R. Al-Balushi, S. A. Al-Araimi, Bearing fault detection using artificial neural networks and genetic algorithm, EURASIP Journal on Applied Signal Processing, 3 (2004), 366-377.

[236] C.S. Tyagi, A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing, World Academy of Science, Engineering and Technology, 43 (2008).

[237] Jayaswal, Pratesh, S.N. Verma, A.K. Wadhwani, Application of ANN, Fuzzy Logic and Wavelet Transform in machine fault diagnosis using vibration signal analysis", Journal of Quality in Maintenance Engineering, 16 (2) (2010), 190-213.

[238] J. Zarei, Induction motors bearing fault detection using pattern recognition techniques, Expert Syst. Appl., 39 (2012), 68-73.

[239] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[240] B. Boser, I. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, New York, 1992.

[241] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995), 273-297.

[242] H. J. Shin, D.-H. Eom, and S.-S. Kim. One-class support vector machines – an application in machine fault detection and classification. Computer & Industrial Engineering, 48 (2005), 395-408.

[243] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural network, augmented by genetic algorithms, Mech. Syst. Signal Process., 16 (2002), 373-390.

[244] B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, Artificial neural network and support vector machine with genetic algorithm for bearing fault detection, Engineering Application of Artificial Intelligence 16 (2003), 657-665.

[245] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process., 18 (3) (2004), 625-644.

[246] B. S. Yang, Han T., and Hwang W.-W. Fault diagnosis of rotating machinery based on multi-class support vector machines. Journal of Mechanical Science and Technology 3 (2005), 846-859.

[247] Z. Zhang, L. Wenzhi, M. Shen, Active learning of support vector machine for fault diagnosis of bearings, Lecture Notes in Computer Science 3973 (2006), 390-395.

[248] Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVM ensemble, Mech. Syst. Signal Process., 21 (2) (2007), 688-705.

[249] Y. Yang, D. Yu, J. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, Measurement, 40 (9-10) 2007, 943-950.

[250] M. Pirra, A. Fasana, Garibaldi, S. Marchesiello, Damage identification and external effects removal for roller bearing diagnostics, In: First European conference of the prognostics and Health management society, Germany, July 3-5, 2012.

[251] X. Zhang, J. Zhou, Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines, Mech. Syst. Signal Process., 41 (1) (2013), 127-140.

[252] A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine, Meccanica, 2014, DOI 10.1007/s11012-014-9968-z.

[253] A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Influence of sifting stop criterion of Empirical Mode Decomposition technique (EMD) on roller bearing fault diagnosis, in Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Springer-Verlag (DEU), Proceedings of the third International Conference on Condition Monitoring of Machinery in Non-Stationary Operations CMMNO 2013.

[254] A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Ensemble empirical mode decomposition (EEMD) and Teager-Kaiser energy operator (TKEO) based damage identification of roller bearings using one-class support vector machine, EWSHM, July 2014.

[255] A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Fault diagnosis of roller bearings using ensemble empirical mode decomposition (EEMD) and support vector machine (SVM), In:

Proceedings of , June 2014.

[256] A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Investigating of a sensitive intelligent method for damage identification of roller bearing under various external conditions, In: Proceedings of Italian association of theoretical and applied mechanics (Aimeta) congress , Torino, 17-20 Sep., 2013.

[257] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process., 19 (2) (2005) 259-270.

[258] G. Rilling, P. Flandrin, P. Goncalves, On empirical mode decomposition and its algorithms, In: Proceedings of IEEE EURASIP Workshop on Nonlinear Signal and Image Processing, Gra do (I), June 2003.

[259] C. Junsheng, Y. Dejie, Y. Yu, Research on the intrinsic mode function (IMF) criterion in EMD method, Mech. Syst. Signal Process., 20 (2006), 817-824.

[260] R.T. Rato, M.D. Ortigueira, A.G. Batista, On the HHT, its problems, and some solutions, Mech. Syst. Signal Process., 22 (2008) 1374–1394.

[261] Q. Xie and B. Xuan, EMD algorithm based on bandwidth and the application on one economic data analysis, In: Proceedings of European Signal Processing Conference, Poland, 2007, 2419-2423.

[262] L. Lin, J. Hongbing, Technical note Signal feature extraction based on an improved EMD method, Measurement, 42 (2009), 796-803.

[263] J. Zhang, R. Yan, R.X. Gao, Z. Feng, Performance enhancement of ensemble empirical mode decomposition, Mech. Syst. Signal Process., 24 (2010), 2104-2123.

[264] W. Guo, P.W. Tse., Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis, In: Prognostics and system health management conference, 2010.

[265] J. Lin, Improved ensemble empirical mode decomposition and its applications to gearbox fault signal processing, International journal of computer and science issues, 2012, 194-199.

[266] H. Jiang, C. Li, H. Li, An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis, Mech. Syst. Signal Process., 36 (2013), 225-239.

[267] M. Zhao, J. Lin, X. Xu, Y. Lei, Tacholess envelope order analysis and its application to fault detection of rolling element bearings with varying speeds, Sensors, 13 (2013), 10856-10875.

[268] M.D. Coats, R.B. Randall, Compensating for speed variation by order tracking with and without a tacho signal, In: Proceedings of VIRM10 – Tenth International Conference on Vibrations in Rotating Machinery, London, United Kingdom, 11–13 September, 2012.